La variabilité des réponses dans les enquêtes statistiques

- Estimation théorique et pratique -

Heinrich Strecker
Universités
de
Tubingue et Munich

Rolf Wiegert
Université
de
Tubingue

avec la collaboration de
Jan Peeters, Luc Anciaux, Elisabeth Draelants
Institut national de statistique
Bruxelles

L'institut national de statistique propose des informations impartiales à un prix abordable.

Les informations sont diffusées conformément à la loi, notamment pour ce qui concerne leur confidentialité.

Nos statistiques couvrent huit domaines :

Généralités	Économie et finances
Territoire et environnement	Agriculture et activités assimilées
Population	Industrie
Société	Services, commerce et transports

Tous droits de traduction, d'adaptation, de reproduction par tous procédés, y compris la photographie et le microfilm sont soumis à autorisation préalable de l'Institut national de statistique.

Toutefois, la citation de courts extraits, à titre explicatif ou justificatif, dans un article, un compte-rendu ou un livre, est autorisée moyennant indication claire et précise de la source.

L'analyse des données de base ainsi que les commentaires n'engagent que la responsabilité de leur(s) auteur(s)

Éditeur responsable : Claude Chéruy

© INSTITUT NATIONAL DE STATISTIQUE *éditeur*

B-1000 Bruxelles
Téléphone + 32 (2) 548.62.11 - Fax + 32 (2) 548.63.67

Préface

Depuis plusieurs années, l'Institut national de statistique peut se réjouir de l'étroite collaboration qui s'est nouée entre l'Unité statistiques agricoles et les Drs Heinrich Strecker, professeur émérite aux universités de Tubingue et de Munich, et Rolf Wiegert.
L'opportunité qu'offrait la relative étroitesse de l'univers agricole belge, conjuguée à la politique d'ouverture pratiquée par les responsables des statistiques agricoles, soutenus en cela par les directions qui se sont succédé, ont permis la réalisation de divers projets statistiques communs.

C'est avec plaisir que je puis présenter ici une nouvelle contribution des Professeurs Strecker et Wiegert, dans laquelle, en se basant sur une enquête-pilote belge et des analyses approfondies des données provenant des recensements agricoles belges, ils ont pu formuler d'importantes avancées théoriques qui ont mené à des développements pratiques et de modélisation en ce qui concerne la variabilité des réponses et l'indice d'inconsistance; ces concepts nouveaux ou approfondis viennent d'ailleurs de faire leur entrée, cette année, dans la 3e mise à jour de la prestigieuse "ENCYCLOPEDIA OF STATISTICAL SCIENCES".

La présente publication a été rendue possible grâce à plusieurs de mes collaborateurs qui, sous la supervision du professeur Strecker, ont spontanément oeuvré à traduire, de manière aussi fidèle que possible, les textes originaux en néerlandais et en français. Je tiens à les en remercier comme je tiens aussi à remercier ceux qui en ont assumé la réalisation.

Le Directeur général,

Claude CHERUY.

Table des matières

Introduction 2

Première partie : Variabilité des réponses et ajustement

I. Composants de l'erreur non due à l'échantillonnage et modèles pour déterminer et réduire celle-ci et, plus particulièrement, l'erreur de réponse 4

II. Enquêtes de contrôle pour établir l'erreur de réponse et sa variabilité dans le cas de caractéristiques non-sensibles 9

A. Généralités 9
B. Modèle d'erreur de réponse dans un sondage aléatoire indéfini 11
C. Modèles pour enquêtes de contrôle 20
1. Modèles pour états (stocks) 21
2. Modèles pour flux (mouvements) 26

III. Variance des réponses : estimation de la valeur de caractéristiques constantes dans le temps. 30

IV. Variance des réponses : estimation de la valeur de caractéristiques non-constantes dans le temps - méthodes d'ajustement dans le temps 39

V. Conclusion de la première partie 54

Seconde partie : Variabilité des réponses et sondages stratifiés

I. Modèle d'enquêtes de contrôle avec stratification 55

A. Introduction 55
B. Estimation de la variabilité des réponses en cas de stratification 57

II. Effet de la variabilité des réponses sur les résultats d'enquêtes stratifiées 58

III. Enquête de contrôle dans le cas de caractéristiques constantes dans le temps. Résultats d'une enquête en Belgique 62

IV. Conclusion de la seconde partie 71

V. Tableaux 72

Conclusion générale 82

Résumé 83

Bibliographie 84

Annexes 88

Introduction

L'Institut national de statistique avait déjà publié, en 1985, dans les *Etudes statistiques* n°75 une étude sur la variance des réponses dans les enquêtes statistiques. Ci-après, nous faisons le point sur l'état actuel de nos recherches.

La variabilité des réponses des déclarations individuelles est d'une grande importance pour la qualité des données provenant d'enquêtes statistiques. La variance des réponses, critère essentiel pour calculer la variabilité des réponses, est estimée à l'aide d'enquêtes répétées.
Comme en Europe, tout comme en Belgique, beaucoup de résultats d'enquête sont obtenus en laissant les déclarants compléter eux-mêmes les questionnaires, nous traiterons également dans cette étude des modèles d'enquête sans recenseurs ni intervieweurs. Dans ce cadre, les enquêtes répétées pour déterminer l'effectif peuvent avoir lieu le même jour de référence ou à différents jours. Des motifs organisationnels font que souvent il faille avoir recours à des jours de référence distincts. Dans ce cas, et plus particulièrement en ce qui concerne les caractéristiques variables dans le temps, il convient d'ajuster les valeurs issues des diverses enquêtes répétées à un jour de référence unique.

La première partie de ce texte traite de l'estimation de la variabilité des réponses et de l'ajustement. **La seconde partie** parle d'un modèle de sondage fréquent, les sondages stratifiés, et de l'évaluation de la variabilité des réponses. Beaucoup de sondages utilisent des échantillonnages stratifiés ou des strates formées d'après des résultats classés d'un point de vue économique ou encore des strates optimales. Quoiqu'il en soit utiliser un sondage stratifié dans le modèle général des enquêtes répétées nécessite d'adapter les formules aux faits. L'adaptation des formules n'est toutefois pas toujours suffisante; en effet, dans les sondages stratifiés les estimations peuvent être influencées par une éventuelle classification erronée. Il faudra aussi estimer l'ampleur de celle-ci.
Ce texte s'appuie sur deux recensements effectués en Belgique, les recensements agricoles et horticoles au 15 mai 1979 (variable "cheptel porcin" dans les exploitations) et au 15 mai 1985 (variable "superficie des étendues cultivées avec cultures" - catégories 1 et 5). Pour le concept ajustement, servent d'exemple les données du recensement de 1979 et celles de la catégorie 5 provenant du recensement de 1985 (cf. partie I).

Le modèle de sondage "Stratification avec variabilité des réponses" est basé sur les résultats de la catégorie 1 du recensement du 15 mai 1985.

Signalons encore une particularité. En de menus détails, l'utilisation des symboles diffère dans les parties I et II au motif que l'on a traité des énoncés différents, ce qui a eu une répercussion sur la symbolisation. Le contexte rend toutefois claire la signification des symboles.

Ce projet est né d'une collaboration entre l'Institut national de statistique et l'ancienne section "statistiques et mathématiques" du séminaire d'économie de l'université de Tubingue en Allemagne.

PREMIERE PARTIE

VARIABILITE DES REPONSES ET AJUSTEMENT

I. Composants de l'erreur non due à l'échantillonnage et modèles pour déterminer et réduire celle-ci et, plus particulièrement, l'erreur de réponse

Les données statistiques primaires (provenant de la statistique publique et privée, des études de marché et d'opinion et des enquêtes sociales) ne s'obtiennent qu'au moyen d'enquêtes exhaustives, de sondages ou de panels. Avant d'entamer une enquête, quelle qu'elle soit, il faut résoudre les problèmes d'adéquation et de conception d'un système de travail.

Les enquêtes, et par là nous entendons communément les méthodes de collecte des données, requièrent la conception d'un système de travail - désigné, en abrégé, par la lettre G (*generalized conditions*) - , c'est-à-dire, en substance, la conversion des concepts et définitions trouvés par adéquation en un schéma d'organisation et de méthode de travail.

Un système de travail est composé comme suit :

a) objectif de l'enquête, devis et calendrier;

b) définition des unités statistiques, des caractéristiques et des modules; établissement des échelles en matière de caractéristiques;

c) programme des tableaux;

d) organisation du travail sur le terrain;

e) s'il y a échantillonnage : le choix du plan de sondage, y compris le choix de la base de sondage, le type - sondage stratifié, sondage à degrés, etc. - , le choix des unités avec sélection à probabilité variable, ainsi que l'estimation ponctuelle et l'intervalle d'estimation;

f) collecte des données, contrôles opérationnels et descriptifs;

g) traitement des données;

h) contrôle de la validité des données, contrôle des erreurs;

i) établissement des tableaux;

j) publication des résultats et diffusion parmi les utilisateurs.

Diverses possibilités s'offrent pour élaborer un système de travail tant au niveau du contenu qu'au niveau opérationnel. C'est ainsi que sur le terrain le déclarant peut compléter un questionnaire sans aide ni assistance, mais on peut aussi imaginer qu'il fournisse lui-même périodiquement les données ou encore que les données soient obtenues au moyen d'interviews. De telles variantes, aussi dans le choix du type d'enquête, mènent à différents systèmes de travail, G, G', G'', etc., bien que l'objectif de l'enquête soit identique. Si les systèmes de travail sont effectués correctement selon les prescriptions, l'on obtient pour la même valeur réelle X les valeurs respectives X_G, $X_{G'}$, $X_{G''}$, etc. Ces valeurs X_G, $X_{G'}$, $X_{G''}$ seront forcément différentes bien qu'elles se rapportent au même X réel. Les écarts entre ceux-ci que l'on peut considérer comme une sorte de variabilité des systèmes de travail, ne peuvent en pratique qu'être à peine définis au motif que, pour des raisons de temps et de coût, l'on ne peut réaliser en principe qu'un seul système de travail.

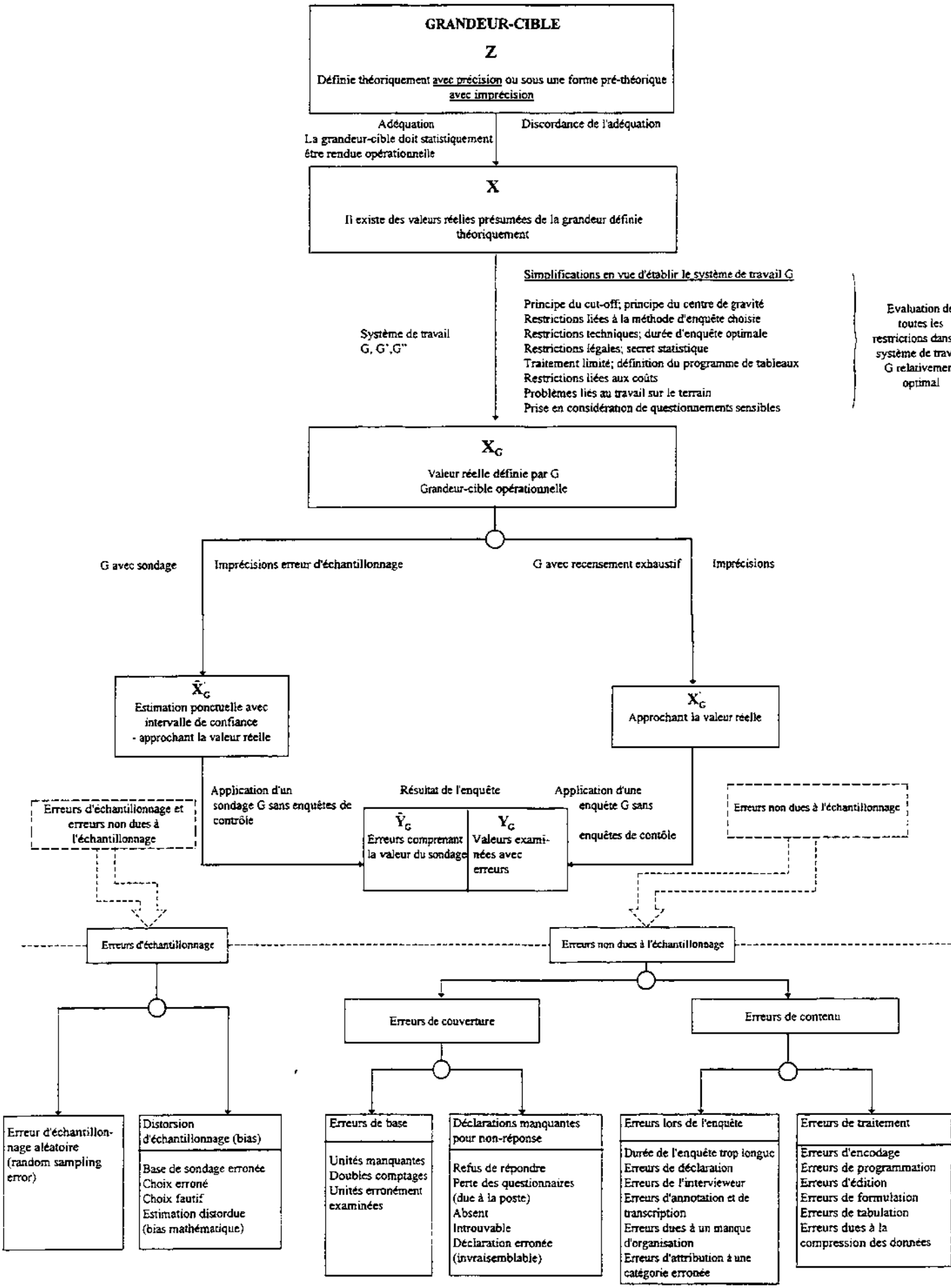
Lors de la collecte des données surgissent des erreurs d'échantillonnage et d'autres non dues à celui-ci (*non-sampling error*). L'influence de l'erreur d'échantillonnage aléatoire (*random sampling error*), qui est un composant essentiel de l'erreur d'échantillonnage, sur l'efficacité des différents modèles de sondage a déjà été amplement étudiée en sorte que poursuivre des recherches ne peut qu'apporter des améliorations marginales. Pour répondre aux exigences accrues que l'on pose à la qualité des données, il vaut mieux s'intéresser à l'erreur non due à l'échantillonnage et à ses divers composants afin que ces erreurs, là où c'est possible, puissent être réduites ou à tout le moins que l'étendue de celles-ci puisse être estimée. Dans le schéma qui suit et qui est basé partiellement sur les travaux de Hurwitz et Pritzker (cf. KANTOROWITZ, 1969) et sur de récents travaux personnels, nous avons représenté pour ce qui est des enquêtes statistiques primaires les méthodes de travail, les simplifications nécessaires à l'opérationalisation et les erreurs que cela entraîne inévitablement. Plus particulièrement, l'erreur d'échantillonnage et l' erreur non due à l'échantillonnage sont manifestement des facteurs importants influençant la qualité des données.

A l'aide de ce schéma, nous pouvons donner quelques éclaircissements sur l'erreur non due à l'échantillonnage.

L'erreur de couverture (*coverage error)* compte deux composants : l'erreur de non-réponse (*non-response error*) et l'erreur due à une base de sondage incomplète (*defect in census coverage error*). Sur la correction de cette dernière erreur et les méthodes à utiliser pour ce faire, voir, entre autres, KISH, 1965; HANSEN, HURWITZ, JABINE, 1963; ZARKOVICH, 1966; LESSLER, KALSBEEK, 1992. Ces dernières années, l'erreur de non-réponse a suscité davantage l'intérêt dans la mesure où l'opinion publique, dans certains pays occidentaux et particulièrement en Allemagne, de par une plus grande sensibilisation aux problèmes de la vie privée, s'est montrée plus critique envers la statistique officielle.

SCHEMA

Interaction entre conception, adéquation, simplification et erreurs dans des enquêtes statistiques primaires



Source : STRECKER, WIEGERT, Wirtschaftsstatistiche Daten und ökonomische Realität, Jahrbücher für Nationalökonomie und Statistik, Band 206 (1989), 490/491.

Cette erreur de non-réponse motivée en partie par des raisons politiques diffère, tant au niveau des causes qu'au niveau du volume, de la non-réponse à des questions sur des sujets sensibles.

Jusqu'à présent l'on n'a pas examiné si les méthodes habituelles, telles les contrôles de probabilité, les enquêtes de contrôle et les méthodes d'échantillons superposés (*interpenetrating samples*), étaient utilisables pour résoudre de tels problèmes de non-réponse.

Afin de réduire l'erreur de non-réponse lorsqu'on aborde des sujets sensibles comme le revenu, la consommation d'alcool, ou encore le casier judiciaire, l'on a développé des modèles de questionnement avec réponses codées au hasard (*randomized response*) (voir WARNER, 1965; CHAUDHURI, MUKERJEE, 1988). L'arrangement au hasard (*randomisation*) permet de rendre anonymes les réponses de manière à ce que l'intervieweur ne puisse voir comment la personne interrogée a répondu à une question sensible. Ainsi sont réduites les réticences à donner des informations sensibles. Il n'est pas encore clair en quelle mesure cette technique peut aussi être utilisée pour réduire l'erreur de non-réponse lors de questionnements politiquement sensibles; il semble toutefois que cette technique onéreuse n'est pratiquement pas applicable lors de recensements et autres grandes enquêtes de même type.

En général, l'effet d'une erreur de non-réponse relativement grande sur la qualité des données statistiques est important et peut à peine être corrigé. Pour réduire cette erreur l'on a par le passé avancé et utilisé divers modèles et méthodes (HANSEN, HURWITZ, 1946; DEMING, 1953; POLITZ, SIMMONS, 1949, 1950; HENDRICKS, 1949; O'MUIRCHEARTAIGH, 1977; GROVES, 1989; BIEMER, GROVES, LYBERG, MATHIOWETZ, SUDMAN, 1991). Il est à craindre qu'à l'avenir, dans un certain nombre de pays occidentaux, l'erreur de non-réponse se développe à cause d'attitudes de rejet pour motifs politiques. D'autres travaux de recherche semblent nécessaires afin de réduire cette influence peu souhaitable ou à tout le moins en mesurer l'étendue.

A côté de l'erreur de couverture il faut signaler l'erreur de contenu (*content error*). Les composants les plus importants en sont l'erreur due au travail sur le terrain et l'erreur de traitement. Le premier comporte principalement des erreurs de réponse et des erreurs liées à l'intervieweur. Celles-ci sont toutefois plus importantes que les erreurs de notation et de transcription que l'on trouve aussi dans le schéma. La méthode des échantillons superposés (*interpenetrating samples*), basée sur les travaux de Mahalonobis (1946), est considérée comme valable et éprouvée. Son but principal est de contrôler et de mesurer l'erreur de l'intervieweur qui influence la façon de répondre et même la réponse (déviation de l'enquêteur).

Dans le texte qui suit, il n'est tenu compte que d'un seul système d'enquête que les pays industrialisés d'Occident et leurs statistiques (publiques) utilisent souvent et dans lequel les personnes interrogées complètent elles-mêmes les questionnaires sans intervention des intervieweurs ou recenseurs.

Cette méthode exclut l'erreur liée à l'intervieweur et fait qu'en conséquence l'erreur liée au travail sur le terrain coïncide quasiment avec l'erreur de déclaration ou de réponse. Pour déterminer en pratique cette méthode l'on a développé et testé des modèles d'enquêtes répétées (voir HANSEN, HURWITZ, PRITZKER, 1961, 1964; FELLEGI, 1964, 1974; BAILAR, DALENIUS, 1969; STRECKER, WIEGERT, PEETERS, KAFKA, 1983).

L'informatique réduit grandement l'erreur de traitement et ses composants mentionnés sur le schéma telles que l'erreur de codage, l'erreur de lecture, etc. L'on peut vérifier partiellement et peut-être même corriger ces erreurs en effectuant des contrôles de vraisemblance sur le matériel collecté. De telles méthodes permettent de corriger aussi bien d'évidentes erreurs de réponse que des erreurs de traitement et plus particulièrement des erreurs de codage. Ont aussi leur place ici les modèles pour corriger automatiquement les déclarations individuelles d'un questionnaire. L'on peut distinguer entre méthodes itératives déterministes et probabilistes.

Comme nous l'avons déjà dit, l'erreur de réponse ou de déclaration est un des composants importants de l'erreur non due à l'échantillonnage d'une méthode d'enquête primaire dont nous parlerons plus loin et dans laquelle il est prévu que les personnes interrogées remplissent elles-mêmes les questionnaires. L'ordre de grandeur de cette erreur de déclaration est un critère pour la validité des données obtenues selon une telle méthode. Ci-après, seront décrits avec leurs principales caractéristiques quelques modèles qui peuvent, si l'on dispose d'états et de flux, être utilisés pour mesurer, en cas de questionnaire complété par la personne interrogée, l'erreur de déclaration individuelle et sa variance.

II. Enquêtes répétées pour établir l'erreur de réponse et sa variabilité dans le cas de caractéristiques non-sensibles

A. Généralités

Dans les pages qui suivent seront évoquées les enquêtes sur des caractéristiques non-sensibles. Les individus ou institutions, interrogés dans un univers ou un échantillon, remplissent le jour de référence les questionnaires qu'on leur a remis sans mode d'emploi et les transmettent par la poste ou par l'entremise d'un recenseur à l'institut de statistique. Dans le cas de caractéristiques non-sensibles et d'une obligation simultanée de déclaration, l'erreur de non-réponse reste en général dans des limites acceptables. Ce n'est pas le cas en ce qui concerne la variabilité des réponses individuelles, qui forme lors d'enquêtes où l'on doit soi-même compléter les formulaires, une importante source d'erreurs et doit pour cette raison être réexaminée à chaque fois et mesurée comme une variance. De cette façon l'on peut définir un composant important de la variance d'erreur totale des données collectées. Pour ce faire, l'on a d'abord développé en théorie des modèles avec enquêtes répétées, qui, par après, ont été élaborés en pratique pour être utilisés avec des caractéristiques tant constantes dans le temps que non-constantes.

Pour déterminer une variance des réponses il faut au moins deux valeurs. Cela signifie qu'il faudrait que le jour de référence aient lieu dans chaque unité interrogée au moins deux enquêtes aussi rapprochées que possible. Dans la pratique statistique des enquêtes et plus particulièrement dans de grands recensements cela n'est guère réalisable. La seule possibilité est de répéter dans un bref laps de temps le dénombrement même ou des pans de celui-ci et de demander à nouveau la valeur de la variable au jour de référence.

Les valeurs d'une variable *non-constante* à différents moments peuvent fortement diverger de sorte qu'il est en fait nécessaire, pour appliquer de manière pertinente des enquêtes répétées à la mesure des réponses, de connaître approximativement la durée réelle des valeurs réelles de la variable et de pouvoir reproduire celle-ci dans un modèle stochastique adéquat pour toutes les unités N de l'univers. Si cela était possible, l'on pourrait convertir des réponses éloignées dans le temps en des réponses multiples au même moment. Comme sur ce sujet l'on n'a que des connaissances très limitées, il faut essayer de résoudre séparément, à l'aide de l'ajustement dit dans le temps, les problèmes, créés par la variabilité des valeurs des variables au sein des modèles d'enquêtes répétées.

Les considérations formelles qui suivent peuvent d'abord servir de base à la modélisation et ensuite à des formes ultérieures, plus pragmatiques de celle-ci tant en ce qui concerne les changements dans le temps de la valeur de la variable qu'en ce qui concerne les erreurs dans les enquêtes mêmes. Prenons une variable X (constante ou pas dans le temps) définie pour N unités statistiques d'un univers approprié à l'enquête. Les valeurs réelles de cette variable sont considérées à $t = 1, 2, \ldots$ moments successifs. Ces valeurs classées dans le temps forment une série chronologique; toutes les séries chronologiques N ensemble sont des valeurs réelles du procès stochastique qui est l'émanation de l'apparition dans le temps des valeurs des variables dans l'univers. Dans les enquêtes sans répétition l'on ne peut toutefois observer le jour de référence qu'une seule valeur réelle pour l'unité. Si l'on pouvait traiter l'univers comme un panel, l'on pourrait en théorie effectuer encore plus d'observations dans le temps. Cette idée n'est que pure hypothèse et dans le cadre d'enquêtes répétées il faut se limiter à deux ou au maximum trois dénombrements dans un délai rapproché.

Les variables économiques sont généralement réparties selon des scores métriques desquels l'on peut déduire l'existence de valeurs réelles qui prennent la forme de valeurs dans une série chronologique x_{it} ($i = 1,\ldots, N$; $t = 1, 2,\ldots$). Dans des enquêtes concrètes ces valeurs réelles ne sont généralement pas définissables. La valeur réelle de la série chronologique t_0, x_{it_0}, que nous représenterons par la suite par x_i, est remplacée selon le système de travail G par une valeur réelle approximative x'_i. Cette grandeur aussi ne peut pas être définie dans la pratique usuelle des enquêtes. Pour le travail statistique ultérieur nous ne disposons que de la valeur y_i observée au moment t_0 et influencée par toutes les erreurs d'enquête, sauf si l'on effectue des examens séparés comme, par exemple, des enquêtes de contrôle pour définir les valeurs réelles x'_i.

Comme nous l'avons déjà signalé, la variabilité individuelle des réponses est un composant significatif de la variance des erreurs totale des données collectées par enquête. Elle ne peut être définie que si pour chaque unité l'on dispose d'au moins deux valeurs observées le même jour de référence t_0. En théorie, il est possible de répéter selon le même système de travail l'observation de la valeur au moment t_0 à différents moments t_1, t_2, ... (dans le temps après t_0); mais en pratique, il faut disjoindre les moments t_0, t_1, t_2, ... de telle manière que dans les intervalles le travail sur le terrain des différentes enquêtes puisse se faire indépendamment. Comme l'étalement des jours de référence en cas de variables non-constantes dans le temps produit en général des changements dans la valeur des variables, les données observées les différents jours de référence t_0, t_1, t_2 doivent être ajustées, c'est-à-dire converties en un seul jour de référence t_0.

Les résultats ajustés (valeurs observées, entachées d'erreurs) des enquêtes répétées sont des grandeurs aléatoires desquelles l'on peut mesurer la variabilité dans leur distribution au moyen d'une estimation de la variance concomitante.

A l'avenir, nous utiliserons la définition suivante : une enquête répétée (recensement répété) au moment t_1 sur l'enquête principale le jour de référence t_0 est une enquête effectuée à un moment postérieur à l'enquête principale avec le *même* système de travail. Lors de l'enquête répétée la variable, tout comme lors de l'enquête principale, est mesurée pour le jour de référence (moment de l'enquête) t_0.

Avant que nous puissions aborder la définition pratique - modèles d'erreurs d'enquête pour recensements répétés - et l'estimation de la variance des réponses et de la variabilité des réponses, nous voulons d'abord exposer, comme base de ce qui suit, le modèle d'erreur de réponse issu de la statistique mathématique dans un sondage aléatoire indéfini (sans intervention de recenseurs ou d'interviewers).

Comme nous l'avons déjà souligné dans la préface, il est difficile d'utiliser les mêmes symboles dans les deux parties de notre étude qui diffèrent par trop de par l'accent mis sur leurs objectifs. Les valeurs de déclaration de l'enquête principale sont symbolisées par (1), celles du premier recensement répété (dénommé enquête-pilote dans l'exemple belge) par (2), celles du second par (3), etc. Par souci de clarté, le jour de référence de l'enquête principale sur des états (stocks) est indiqué par t_0, celui du premier recensement un autre jour de référence par t_1, celui du second par t_2, etc. Il en va de même, toute proportion gardée, pour les enquêtes sur des flux (mouvements).

B. Modèle d'erreur de réponse dans un sondage aléatoire indéfini

Soit une population au moment précis t_0, composée de $i = 1,\ldots, N$ unités, porteuses d'une caractéristique X. Au moment t_0 une valeur x_i (valeur réelle) existe pour chaque unité i. L'on désire ensuite collecter statistiquement cette valeur réelle x_i au moyen d'une méthode appropriée (système de travail d'une enquête avec questionnement individuel). Selon la façon dont a été conçu le système de travail, l'observation (collecte des réponses individuelles) peut subir des influences menant à des erreurs systématiques et aléatoires; cela donne pour chaque unité une valeur observée y_i. Le modèle théorique de l'enquête répétée prescrit alors une réitération $t = 1, \ldots, k, \ldots$ simultanée de l'observation de la valeur réelle x_i afin de pouvoir traiter statistiquement par observation répétée l'action des erreurs.

Cela conduit à des valeurs observées (réponses) $y_{i1}, y_{i2}, \ldots, y_{it}, \ldots$ (*trials*) présentant une certaine variabilité. Pour la réponse individuelle y_{it} l'on a alors la représentation et la décomposition suivantes :

$$y_{it} = \underbrace{x_i}_{\text{valeur réelle}} + \underbrace{(y_i - x_i)}_{\text{erreur systématique } e_i} + \underbrace{(y_{it} - y_i)}_{\text{erreur aléatoire } \varepsilon it}$$

où $e_i = E_t(e_{it})$ et $e_{it} = y_{it} - x_i$ sont l'erreur de réponse ou de déclaration individuelle. Il résulte de la décomposition de y_{it} que la variance des réponses individuelles

$$\sigma^2_{R_i} = E_t\left[(y_{it} - y_i)^2 | i\right] = E_t\left[\varepsilon_{it} - \underbrace{E_t(\varepsilon_{it})}_{0}\right]^2 = E_t(\varepsilon_{it}{}^2)$$

est la variance individuelle de l'erreur aléatoire ε_{it}.

Cette représentation de base d'un processus de réponse stochastique peut schématiquement, en ce qui concerne le moment t_0, être représentée comme suit. Pour un moment t_1 (situé dans le temps avant ou après t_0), il existe un schéma analogue. Si la variable X dépend du temps, les valeurs de t_1 seront différentes de celles de t_0; par analogie cela est aussi valable pour t_2, etc. Il en résulte pour l'utilisation pratique des enquêtes répétées que celles-ci, par la méthode du soi-disant ajustement, doivent être converties à différents moments au même jour de référence défini théoriquement; en d'autres mots, le schéma de t_1 ou t_2, etc. doit être ramené à celui de t_0. Dans la présente partie, nous traiterons de cette problématique de façon détaillée dans les chapitres III et IV pour ce qui est des caractéristiques temporellement dépendantes et constantes.

Schéma

Unité n°.	X	observations répétées	espérance	variance des réponses individuelles
1	x_1	y_{11}.....y_{1t}.....	.	.
2	x_2	y_{21}.....y_{2t}.....	.	.
.	.		.	.
.	.		.	.
.	.		.	.
i	x_i	y_{i1}.....y_{it}.....	$E_t(y_{it}\|i) = y_i$	$\sigma^2_{R_i} = E_t[(y_{it} - y_i)^2 \| i]$
.	.			.
.	.			.
.	.		.	.
N	x_N	y_{N1}.....y_{Nt}.....		.

Formules des paramètres de l'univers :

$$\overline{X} = \overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} x_i ; \overline{Y} = \overline{Y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i ;$$

$$\sigma_y^2 = \underset{i}{E}(y_i - E(y_i))^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y}_N)^2$$

$$\sigma_R^2 = \underset{i}{E}(\sigma_{Ri}^2) = \frac{1}{N}\sum_{i=1}^{N}\sigma_{Ri}^2$$

σ_R^2 est la variance des réponses totale et en même temps la variance totale de toutes les erreurs aléatoires ε_{it} dans les valeurs données.

Comme l'on ne peut réaliser en pratique un "nombre infini" de recensements de contrôle t = 1, 2, 3 ....., k, ... , l'on n'aura que des formules estimatives concernant un nombre limité de recensements (répétés) t = 1, 2, 3 ....., k, donnant au niveau de l'univers, de l'échantillon, de la valeur moyenne et de la valeur totale :

	Univers	Echantillon	
Valeur moyenne	$\hat{\overline{Y}}_{(K)N} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{k}\sum_{t=1}^{K} y_{it}$;	$\overline{y}_{(K)n} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{k}\sum_{t=1}^{K} y_{it}$.	n = taille de l'échantillon n < N
Valeur totale	$\hat{Y}_{(K)N} = \sum_{i=1}^{N}\frac{1}{k}\sum_{t=1}^{K} y_{it}$;	$\hat{Y}_{(K)n} = \frac{N}{n}\sum_{i=1}^{n}\frac{1}{k}\sum_{t=1}^{K} y_{it}$.	k = nombre d'enquêtes répétées selon le même système de travail

Tout comme dans l'analyse de séries chronologiques, l'objectif dans les enquêtes doit être de définir de manière numérique les composants particuliers tels, entre autres, la moyenne des valeurs réelles $\overline{x}$,l'erreur systématique $\overline{e}$, l'erreur aléatoire $\overline{\varepsilon}$. L'intervalle de variation de l'erreur aléatoire ε_{it} dans les résultats d'enquête est mesuré par l'expression mathématique de la variabilité des réponses à l'aide de la variance des réponses. Avec un nombre croissant de recensements répétés le chiffre absolu de la valeur moyenne de l'erreur aléatoire $\left|\hat{\overline{\varepsilon}}_{(K)N}\right|, \left|\hat{\overline{\varepsilon}}_{(K)n}\right|$ se réduit sans cesse d'un point de vue stochastique et dans le cas d'un "nombre infini" de recensements répétés devient

$$\left|\hat{\bar{\varepsilon}}_{(K=\infty)N}\right| = 0, \quad \left|\hat{\bar{\varepsilon}}_{(K=\infty)n}\right| = 0$$

ou énoncé mathématiquement

$$\operatorname{plim}_{k\to\infty} \hat{\bar{\varepsilon}}_{(k)N} = 0, \qquad \operatorname{plim}_{k\to\infty} \hat{\bar{\varepsilon}}_{(k)n} = 0,$$

L'on peut en conclure que dans le cas d'un nombre pas trop restreint de recensements répétés k et d'une valeur moyenne (approximative) des valeurs réelles : $\overline{X}_N, \bar{x}_n$, l'erreur systématique moyenne d'une enquête (moyenne de l'erreur systématique individuelle) est respectivement d'environ $\bar{e}_N \approx \hat{\bar{Y}}_{(K)N} - \overline{X}_N$ et $\bar{e}_n \approx \bar{y}_{(K)n} - \bar{x}_n$ et qu'elle peut être définie numériquement. Comme concrètement il est difficile d'effectuer plus de deux enquêtes en même temps, il est en général impossible, lorsqu'il n'y a que K=2 recensements (une enquête principale et un seul recensement répété), de donner une définition précise de l'erreur systématique d'une enquête, soit que la valeur individuelle observée y_{it} ne comprend pas d'erreur aléatoire ε_{it}, soit que pour définir les composants d'erreur systématiques et aléatoires l'on introduise une hypothèse appropriée plausible. L'on peut la plupart du temps définir par approximation les valeurs individuelles réelles x_i au moyen de valeurs de contrôle provenant de l'enquête de contrôle, pour autant qu'une fois de plus le comptage ait été effectué "correctement", selon le même système de travail, sous forme de sondage ou de sous-échantillon.

Pour ne pas trop surcharger les personnes interrogées - rappelons qu'il s'agit d'un modèle d'enquête dans lequel l'on remplit soi-même le questionnaire, sans "erreur liée à l'intervieweur" - il est recommandé dans le cas d'une enquête principale comme recensement exhaustif de réaliser le recensement répété comme sondage, et dans le cas d'une enquête principale comme sondage l'enquête répétée comme sous-échantillon. En conséquence, il faut abandonner les formules définissant les valeurs estimées de sondages et les remplacer par :

Formules des valeurs estimées des échantillons :

n taille de l'échantillon, $n \leq N$

k nombre d'enquêtes répétées selon le même système de travail; $t = 1,2,...k$

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \underset{t}{E}(y_{it}|i) \; ; V(\bar{y}_n) = \underset{i}{E}(\bar{y}_n - \underset{i}{E}(\bar{y}_n))^2$$

$$\bar{y}_{(k)n} = \frac{1}{n \cdot k}\sum_{i=1}^{n}\sum_{t=1}^{k} y_{it} \; ; \; V(\bar{y}_{(k)n}) = \underset{i,t}{E}(\bar{y}_{(k)n} - \underset{i,t}{E}(\bar{y}_{(k)n}))^2 \; ; \; \bar{Y}_N = \underset{i,t}{E}(\bar{y}_{(k)n})$$

$$s_{Ri}^2 = \frac{1}{k-1}\sum_{t=1}^{k}(y_{it} - \frac{1}{k}\sum_{t=1}^{k} y_{it})^2$$

$$s_R^2 = \frac{1}{n}\cdot\sum_{i=1}^{n} s_{Ri}^2$$

estimation de la variance des réponses individuelles et totales

$$\hat{V}_R = \frac{s_R^2}{n \cdot k}$$

estimation d'une variabilité des réponses simplifiée sans tenir compte des corrélations $\rho_t, \rho_R, \rho_R^{'}$

Erreur Quadratique Moyenne (EQM) :

(*)
$$E\hat{Q}M(\bar{y}_{(k)n}) = \underset{i,t}{E}[\bar{y}_{(k)n} - \bar{X}_N]^2 = \underbrace{\underset{i,t}{E}(\bar{y}_{(k)n} - \bar{Y}_N)^2}_{V(\bar{y}_{(k)n})} + B^2$$

et biais $B = \bar{Y}_N - \bar{X}_N$.

L'EQM définie comme telle peut être analysée comme suit :

(**)
$$E\hat{Q}M(\bar{y}_{(k)n}) \approx \underbrace{\frac{\sigma_R^2}{n \cdot k}[1 + (k-1)\rho_t + (n-1)\rho_R + (n-1)(k-1)\hat{\rho}'_R]}_{\text{variabilité des réponses}} + V(\bar{y}_n) + B^2$$

ρ_t, ρ_R, ρ'_R sont définis comme:

auto-corrélation (complète)

$$\rho_t = \frac{\underset{i,t \neq t'}{E}(y_{it} - y_i)(y_{it'} - y_i)}{[\underset{i,t}{E}(y_{it} - y_i)^2 \cdot \underset{i,t'}{E}(y_{it'} - y_i)^2]^{1/2}} \qquad t \neq t'$$

corrélation de réponses provenant d'unités différentes $i \neq j$ dans une même enquête t

$$\rho_R = \frac{\underset{i \neq j,t}{E}(y_{it} - y_i)(y_{jt} - y_j)}{[\underset{i,t}{E}(y_{it} - y_i)^2 \cdot \underset{j,t}{E}(y_{jt} - y_j)^2]^{1/2}}$$

corrélation de différentes unités $i \neq j$ et de différentes enquêtes $t \neq t'$

$$\rho'_R = \frac{\underset{i \neq j,t \neq t'}{E}(y_{it} - y_i)(y_{jt'} - y_j)}{[\underset{i,t}{E}(y_{it} - y_i)^2 \cdot \underset{j,t'}{E}(y_{jt'} - y_j)^2]^{1/2}}$$

En comparant les formules (*) et (**) de l'EQM, nous obtenons :

$$V(\bar{y}_{(k)n}) \approx \frac{\sigma_R^2}{n \cdot k}[1 + (k-1)\rho_t + (n-1)\rho_R + (n-1)(k-1)\rho_R'] + V(\bar{y}_n).$$

Un tel modèle d'enquête nécessite en pratique des moyens humains et financiers considérables et représente une lourde charge pour les personnes interviewées. Pour ces raisons il n'est guère facile de l'organiser. L'on comprendra aisément qu'en pratique, en plus de l'enquête principale (enquête originelle), il ne pourra y avoir qu'*une seule* enquête complémentaire, habituellement sous la forme d'un sondage d'une envergure bien moindre. Les modèles d'enquête répétée que l'on utilise dans la pratique sont, dès lors, généralement basés sur une enquête principale et *une* enquête répétée, en d'autres mots, sur un total $k = 2$ enquêtes.

Par cette raison nous reproduisons ici séparément les formules du cas particulier $k = 2$ avec les réponses y_{i1}, y_{i2} :

$$\bar{y}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{y_{i1} + y_{i2}}{2} = \frac{1}{n} \sum_{i=1}^{n} \hat{y}_i \qquad \text{(avec replacement)}$$

$$\hat{V}(\bar{y}_n) = \frac{1}{n} \cdot \frac{1}{n-1} \sum_{i=1}^{n} [\frac{y_{i1} + y_{i2}}{2} - \bar{y}_n]^2 = s_y^2 / n$$

En cas de sélection sans replacement il faut multiplier s_y^2 / n par le facteur $(N - n)/N$.

Comme valeur estimée s_R^2 pour σ_R^2 et la variabilité, l'on obtient

$$s_R^2 = \frac{1}{n} \sum_{i=1}^{n} s_{Ri}^2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{2} (y_{it} - \frac{y_{i1} + y_{i2}}{2})^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} (y_{i1} - y_{i2})^2 = \frac{1}{2n} \sum_{i=1}^{n} (y_{i1} - y_{i2})^2$$

$\hat{V}_R = \frac{s_R^2}{2n}$ Valeur estimée de la variabilité des réponses simple

$$E\hat{Q}M(\bar{y}_{(2)n}) = \frac{s_R^2}{2n}[1 + \hat{\rho}_t + (n-1)\hat{\rho}_R + (n-1)\hat{\rho}_R'] + \hat{V}(\bar{y}_n) + \hat{B}^2$$

Pour $k = 2$ l'on peut en déduire de manière triviale $\hat{\rho}_t, \hat{\rho}_R, \hat{\rho}'_R$ si :

$$\hat{\rho}_t = \frac{\sum_{i=1}^{n}(y_{i1} - \hat{y}_i)(y_{i2} - \hat{y}_i)}{[\sum_{i=1}^{n}(y_{i1} - \hat{y}_i)^2 \cdot \sum_{i=1}^{n}(y_{i2} - \hat{y}_i)^2]^{1/2}} = \frac{-\sum_{i=1}^{n}(y_{i1} - y_{i2})^2}{\sum_{i=1}^{n}(y_{i1} - y_{i2})^2} = -1$$

(limite inférieure de la valeur de l'auto-corrélation)

De manière analogue l'on obtient :

$$\hat{\rho}_R = \frac{\sum_{i,j=1;i\neq j}^{n}(y_{i1}-y_{i2})(y_{j1}-y_{j2})}{\sum_{i=1}^{n}(y_{i1}-y_{i2})^2}$$

$$\hat{\rho}'_R = -\hat{\rho}_R.$$

A l'aide de ces valeurs l'on peut donner une estimation de $l'\hat{EQM}$. Si pour ρ_t on substitue la plus grande valeur possible 1 et qu'ensuite on mette du côté droit $\hat{\rho}_R = -\hat{\rho}'_R$, il s'ensuit l'estimation suivante :

$$0 \leq \hat{EQM}(\bar{y}_{(2)n}) \leq \frac{s_R^2}{2 \cdot n}[2 + (n-1)\hat{\rho}_R - (n-1)\hat{\rho}_R] + \hat{V}(\bar{y}_n) + B^2,$$

$$0 \leq \hat{EQM}(\bar{y}_{(2)n}) \leq \frac{s_R^2}{n} + \hat{V}(\bar{y}_n) + B^2.$$

Si l'on prend la plus petite valeur possible -1 pour ρ_t et qu'on met $\hat{\rho}_R = -\hat{\rho}'_R$ dans $l'\hat{EQM}$, nous trouvons pour $k = 2$ la valeur la plus faible suivante :

$$0 \leq \hat{V}(\bar{y}_n) + B^2 \leq \hat{EQM}(\bar{y}_{(2)n}).$$

De la sorte l'on obtient l'estimation suivante pour un estimateur d'Erreur Quadratique Moyenne ($k = 2$):

$$\underbrace{\hat{V}(\bar{y}_n) + B^2}_{} \leq \hat{EQM}(\bar{y}_{(2)n}) \leq \underbrace{\frac{s_R^2}{n} + \hat{V}(\bar{y}_n) + B^2}_{}$$

Minimum estimé de l'erreur quadratique moyenne sans tenir compte de la variance des réponses ($\hat{EQM}_{\min}$)

Maximum estimé de l'erreur quadratique moyenne en tenant compte de la variance des réponses ($\hat{EQM}_{\max}$)

Pour un biais connu ou pouvant être estimé ($\hat{B}$), l'on peut indiquer les valeurs minimales et maximales. Nous trouvons alors :

$$E\hat{Q}M(\overline{y}_{(2)n}) = \frac{s_R^2}{2n} + \hat{V}(\overline{y}_n) + \hat{B}^2$$

(Estimateur d'erreur quadratique moyenne comme estimation ponctuelle)

La limite inférieure de l'estimation est une valeur de l'erreur quadratique moyenne sans tenir compte de la variance des réponses ($s_R^2 = 0$). L'estimation à droite représente la limite supérieure de l'influence de la variance des réponses sur l'estimateur EQM. Dès lors, la part de la variance des réponses (variabilité des réponses) dans l'estimateur EQM peut être estimée comme suit :

$$Q_{Min} := \frac{0}{\sqrt{E\hat{Q}M_{Max}}} \leq Q := \frac{s_R/\sqrt{2n}}{\sqrt{E\hat{Q}M}} \leq Q_{Max} := \frac{s_R/\sqrt{n}}{\sqrt{E\hat{Q}M_{Min}}}.$$

L'importance de cette estimation pour l'EQM (k = 2) réside dans le fait qu'en ce cas - important en pratique - un calcul de l'estimateur pour k = 2 permet de tenir compte de l'influence minimale et maximale des corrélations. Ceci représente un apport efficace à l'estimation ponctuelle. Ce contexte est exposé de façon détaillée dans des travaux antérieurs (STRECKER, 1980; STRECKER, WIEGERT, PEETERS, etc., 1983; STRECKER, PEETERS, KAFKA, WIEGERT, 1985; BECKMANN, WIEGERT, 1987).

C. Modèles pour enquêtes répétées

Cette version théorique du modèle pour enquêtes répétées est aussi utilisable en pratique lorsque la valeur de la variable au moment t_0 a une grande valeur de rappel pour la personne interrogée. Par là, nous entendons que celle-ci se souvient toujours de la valeur t_0, lorsqu'il s'agit de sa date de naissance, de son domicile, de la forme juridique de son exploitation, etc. Si l'on observe une telle variable dans le temps, elle apparaîtra généralement comme (une) constante ou à tout le moins constante sur de longues périodes. Grâce à la grande valeur de rappel des variables il est possible d'effectuer une enquête répétée à des moments t_1, t_2 ,..., qui viennent après t_0, de demander à cette occasion la valeur de la variable au moment t_0 et d'établir ainsi une variabilité des réponses; dans le cas de telles variables celle-ci sera extrêmement faible. Dans cette variante du modèle il n'est pas nécessaire d'effectuer un ajustement comme décrit dans la première partie, chapitre IV.

Si la valeur de la variable au moment t_0 n'a qu'une valeur de rappel réduite, c'est-à-dire si la personne interrogée au moment t_1 ne peut répéter sa déclaration du jour de référence t_0 avec la même acuité mnémonique, elle essayera de reconstituer la valeur au moment t_0 d'après la méthode qu'elle a utilisée dans sa déclaration pour le moment t_0. Cela n'est pas grave si la valeur de la variable durant la période écoulée est restée égale (série chronologique constante), mais ce ne l'est plus si les valeurs de la variable sont fonction du temps (série chronologique non-constante). Avec des grandeurs économiques nous avons généralement affaire à des séries chronologiques non-constantes de sorte que la mensuration de la variabilité des réponses ne peut s'appuyer sur ce modèle théorique d'enquêtes répétées. Tant pour ce motif que pour des raisons de faisabilité pratique des enquêtes il est nécessaire d'apporter des modifications pragmatiques au modèle des recensements répétés. Bien que l'enquête principale le jour de référence t_0 et les enquêtes répétées aux moments t_1, t_2 ,... aient lieu selon le même système de travail, l'on ne demande pas aux moments t_1, t_2 ,... la valeur t_0 mais les valeurs t_1, t_2 ,...de la variable qui avec le temps se sont modifiées. Si l'on observe la série chronologique correspondante, l'on trouvera les termes des valeurs de la variable à différents moments.

Comment peut-on avec ces faits en arriver à établir une variabilité des réponses lorsque des valeurs divergentes d'une série chronologique sont demandées à des moments différents ? Seulement au moyen d'informations complémentaires concernant les données et d'hypothèses additionnelles par lesquelles l'on peut établir une connexion entre les valeurs à différents moments.

L'altération chronologique des variables peut ainsi être isolée et corrigée par un procédé d'ajustement. Ceci est d'autant plus simple dans le cas d'une autocorrélation élevée des valeurs de la série chronologique ou lorsque l'on peut calculer l'estimation d'un patron de remplacement moyen.

Pour des raisons pratiques il faut un certain espacement entre l'enquête principale et les enquêtes répétées. Afin de limiter les coûts et le travail organisationnel et de ne pas poser d'exigences trop élevées à l'"éthique sociale" des personnes interrogées, il est en pratique nécessaire de s'en tenir à des modèles n'offrant qu'*un seul* recensement répété. Nonobstant cette restriction il est possible de définir par approximation la variance des réponses comme un composant essentiel de l'erreur d'enquête totale. Ci-après, seront commentés, répartis d'après leurs caractéristiques, des modèles pour enquêtes répétées selon les états ou les flux.

1. Modèles pour états (stocks)

Précisons d'abord les principaux symboles utilisés pour ces modèles :

x_i	valeur réelle au moment t_0
y_{i1}	valeur observée dans l'enquête principale au moment t_0
$e_{i1} = y_{i1} - x_i$	erreur individuelle de réponse (*respondent error*) dans l'enquête principale
y_{i2}	valeur observée dans l'enquête répétée au moment $t_1 = t_0$ (enquête répétée réelle) ou du jour de référence t_1 vers t_0 valeur ajustée
$e_{i2} = y_{i2} - x_i$	erreur de réponse individuelle dans l'enquête répétée
y_{i2}^{*}	valeur observée dans l'enquête répétée au moment t_1 (valeur non ajustée)
Δ_i	correction (additive) d'ajustement pour convertir la valeur y_{i2}^{*} en valeur y_{i2}

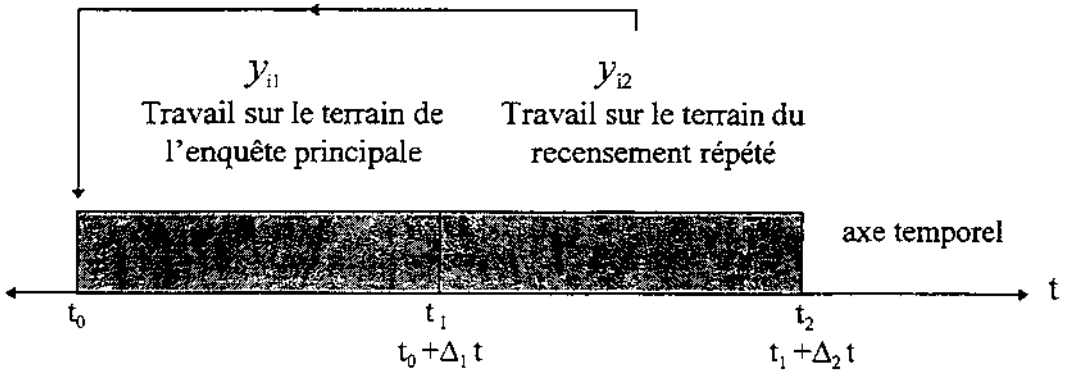


Schéma du modèle (a)

a) L'on effectue une enquête principale au jour de référence t_0 et une enquête répétée dans la période de t_1 à t_2 ; t_1 et t_2 étant situés après la période durant laquelle a lieu le travail sur le terrain de l'enquête principale. Lors du recensement répété l'on demande la valeur de la variable le jour du recensement t_0 (recensement répété réel). L'on obtient comme résultat y_{i1} , y_{i2} . Ceci n'est pas problématique (dans le cas d'une valeur constante ou approximativement constante de la variable). Il est important de savoir s'il s'agit d'une valeur de variable à valeur de rappel élevée ou faible.

b) Ce modèle est utilisé avec des valeurs de la variable à faible valeur de rappel qu'elles soient constantes ou non constantes dans le temps. Comme sous (a) l'enquête principale a lieu au jour de référence t_0 et le recensement répété en dehors de la période du travail sur le terrain et après l'enquête principale. La différence par rapport à (a) est que lors du recensement répété l'on ne demande pas la valeur de la variable au jour de référence t_0, mais la valeur de la variable au moment t_1 (recensement répété ex post comme succédané d'un vrai recensement de contrôle); voir schéma du modèle (b). Les valeurs observées sont indiquées par y_{i1} et y^{*}_{i2}.

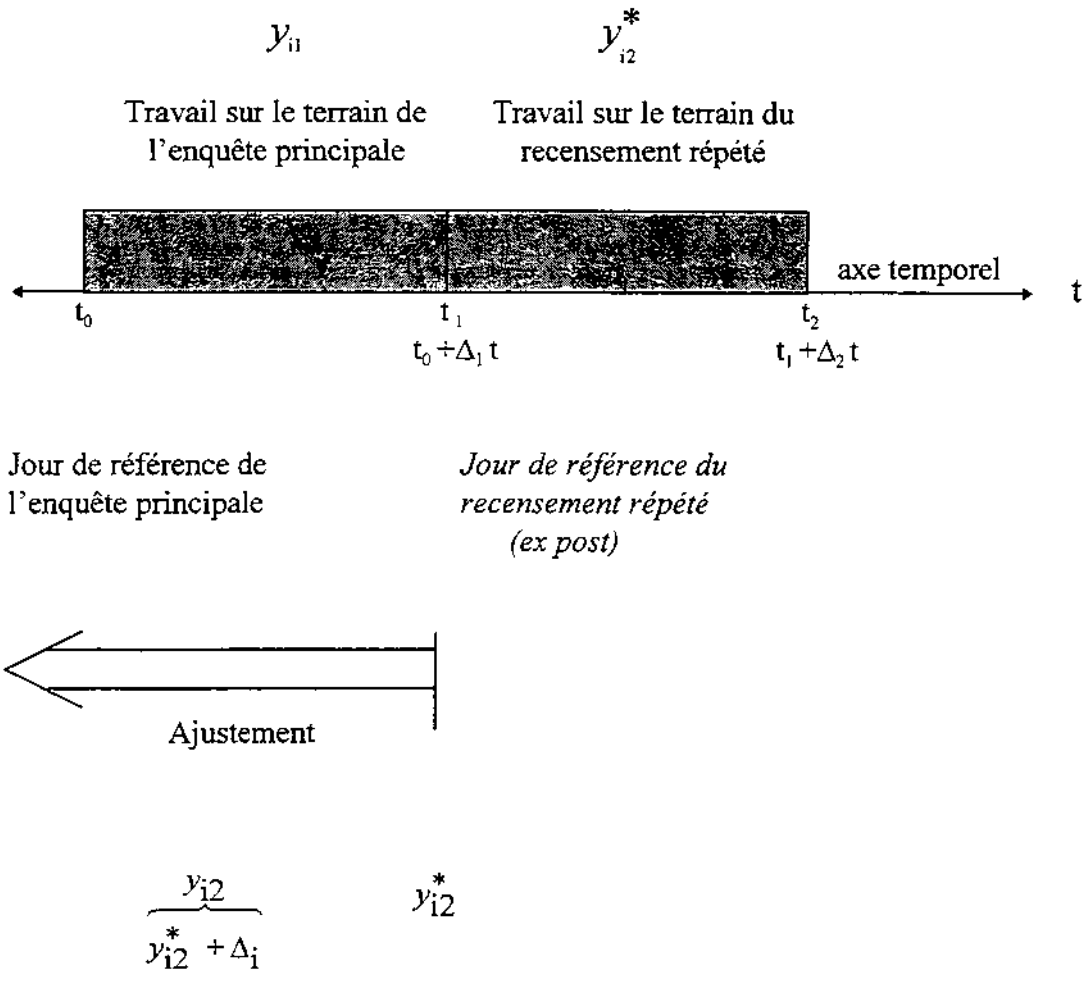


Schéma du modèle (b)

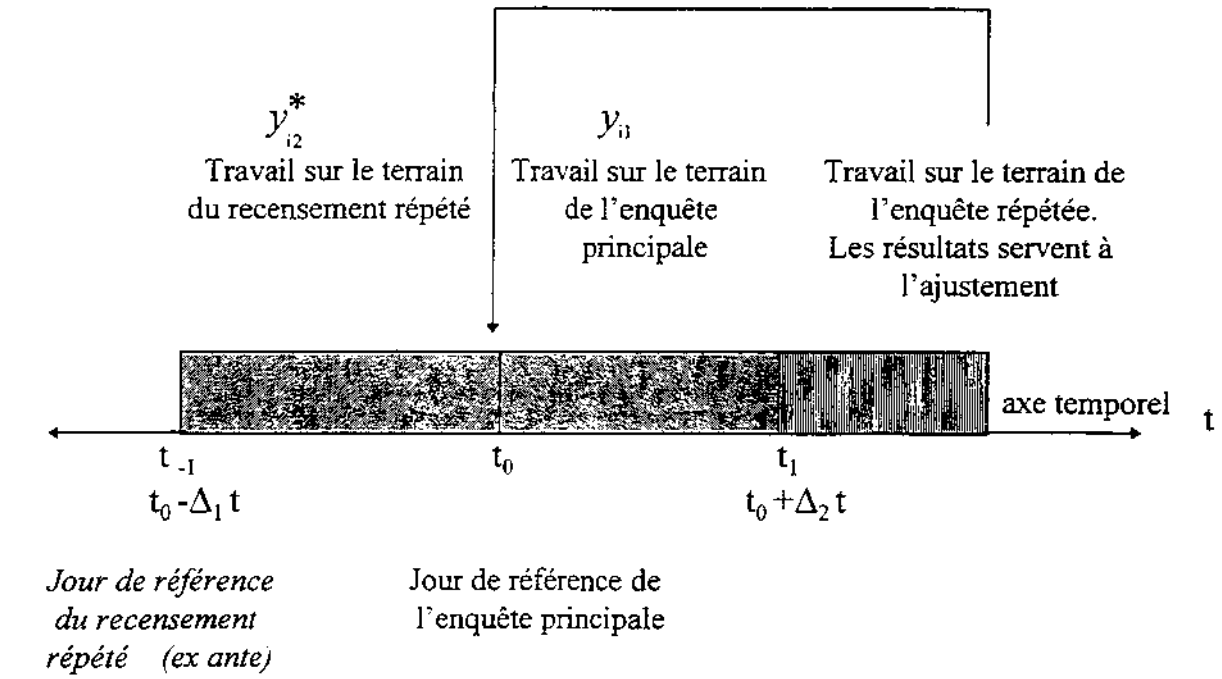
A l'intérieur du modèle (b) il y a deux variantes :

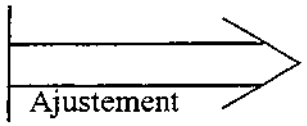
b1) La valeur de la variable est (approximativement) constante dans le temps pour la période $\Delta_1 t = t_1 - t_0$. En conséquence, la valeur observée au moment t_1 - pour autant que soit utilisé le même système de travail - peut tout comme dans le modèle (a) être considérée comme un vrai recensement répété pour l'enquête principale, en d'autres mots $y^*_{i2} = y_{i2}$. Que l'on songe à d'autres exemples tels que l'enquête sur des données personnelles dans la statistique de la population ou encore l'enquête sur les surfaces cultivables dans une période précise de végétation.

b2) La valeur de la variable dépend du temps, en d'autres mots, au jour de référence t_1 la valeur de la variable - que l'on désire observer avec le même système de travail - n'est plus la même. Pour pouvoir quand même considérer l'enquête au jour de référence t_1 comme un recensement répété, il faut supposer qu'il y ait un lien systématique (avec, par exemple, une corrélation élevée) entre les deux valeurs observées, considérées comme valeurs de la variable dans une série chronologique. Une analyse de la dépendance temporelle (analyse de la structure de la série chronologique) permet d'examiner la corrélation interne des valeurs observées.

Un ajustement, c'est-à-dire un "calcul en arrière" de la valeur y^*_{i2} observée le jour de référence t_1 vers la valeur y_{i2} du jour de référence t_0, est nécessaire. La forme spécifique de ce procédé d'ajustement dépend de la structure de la série chronologique qui en est à la base. Le procédé d'ajustement lui-même est donc, dans le cadre d'une réduction sensible de la valeur $\Delta_1 t = t_1 - t_0$ le maillon à l'aide duquel deux enquêtes à des jours de référence t_0,t_1 différents peuvent être liées comme enquêtes répétées. En principe, il existe à l'intérieur de ce modèle (b2) des formes très diverses d'ajustement, par exemple au moyen de procédés basés sur des séries chronologiques ou d'enquêtes complémentaires (établissement des réelles modifications dans le temps à l'aide d'une enquête de contrôle sur place).

Si l'on adapte au modèle (b2) un ajustement particulier, l'on obtient le modèle (c) qui à l'aide d'une enquête complémentaire (enquête de contrôle descriptive) fonctionne comme un sondage. L'enquête de contrôle a pour objectif principal de collecter des informations (valeurs moyennes) sur les erreurs de réponse et de déclaration et sur les vraies modifications dans le temps des valeurs de la variable. En tant que contrôle elle a lieu après l'enquête principale (ex post). Il y a donc deux enquêtes complémentaires (ou bien un recensement répété ex post et après une enquête de contrôle descriptive, ou bien d'abord une enquête de contrôle descriptive et ensuite un recensement répété ex post), à chaque fois à un certain intervalle après l'enquête principale. Les périodes durant lesquelles a lieu le travail sur le terrain des différentes enquêtes ne peuvent se chevaucher; de ce fait il y a généralement un écart assez grand entre l'enquête principale, le recensement répété et l'enquête de contrôle, ce qui peut avoir des effets négatifs.


y^*_{i2}
Travail sur le terrain du recensement répété
y_{i1}
Travail sur le terrain de l'enquête principale
Travail sur le terrain de l'enquête répétée. Les résultats servent à l'ajustement
axe temporel
t
t_{-1}
t_0
t_1
$t_0 - \Delta_1 t$
$t_0 + \Delta_2 t$
Jour de référence du recensement répété (ex ante)
Jour de référence de l'enquête principale



y^*_{i2} $\underbrace{\qquad y_{i2} \qquad}_{y^*_{i2} + \Delta_i}$

Schéma du modèle (c)

c) Pour cette raison le jour de référence t_1 du recensement répété est placé *avant* le jour de réfé-rence de l'enquête principale t_0 (enquête répétée ex ante comme succédané d'une véritable enquête répétée). La durée totale des trois enquêtes reste de la sorte identique, les enquêtes complémen-taires étant disposées de manière relativement symétrique autour du jour de référence t_0 de l'enquête principale. Cela représente un gros avantage en matière d'organisation pratique du travail sur le terrain puisque les intervalles plus courts entre les jours de référence des différentes enquêtes facilitent l'ajustement. Dans ce modèle, t_1 reçoit le nom t_1. Si la valeur de la variable dans la période $\Delta_i t = t_0 - t\_1$ ne change pas (valeur constante dans le temps), il n'est pas important si le recensement répété a lieu avant ou après t_0. Si la valeur de la variable n'est pas constante entre t_1 et t_0, la valeur y^*_{i2} observée au moment t_1 doit être convertie en une valeur y_{i2} au moment t_0 (ajustement).

En principe, cela n'a pas d'importance s'il faut ajuster une valeur antérieure ou postérieure. Dans le schéma du modèle (c) l'on voit comment dans le temps les enquêtes se rapportent les unes par rapport aux autres.

Le modèle (c) est un modèle utilisable. L'on peut dans un laps de temps acceptable organiser les trois enquêtes qui sont nécessaires à la définition de la variabilité des réponses. Grâce à la valeur provenant de l'enquête principale y_{i1} et à la valeur ajustée y_{i2}, l'on obtient une deuxième valeur observée (à tout le moins approchante) pour le jour de référence t_0, à l'aide de laquelle l'on peut calculer la variance des réponses individuelle ($\sigma^2{}_{R_i}$) d'une unité i (porteur de la caractéristique).

2. Modèles pour flux (mouvements)

Les modèles desquels l'on a parlé jusqu'à présent traitent de problèmes qui apparaissent lors d'enquêtes répétées de stocks (nombre d'habitants, total des cultures, nombre d'animaux, etc.). Pour pouvoir utiliser ce modèle lors d'enquêtes sur des flux (nombre de naissances, nombre de mariages, nombre de candidats reçus à un examen, etc., pendant une période de référence), il ne faut pas le modifier fondamentalement, seules de légères adaptations du modèle (a) étant nécessaires. Des recensements répétés concernant des flux ne sont en ce sens qu'un cas particulier des modèles précités.

L'enquête sur des unités de flux et sur la présence de la caractéristique ad hoc s'effectue ordinairement à un moment t_0 à l'issue de la période référence $[\tau_1, \tau_2], \Delta\tau = \tau_2 - \tau_1$. t_0 est souvent égal à τ_2 ou le suit de près dans le temps. Comme la valeur de la variable est constante, il n'est pas nécessaire d'associer à un jour de référence unique des valeurs qui ont été demandées à plus d'une reprise et de les ajuster comme dans le cas de flux. La représentation théorique de l'enquête selon laquelle la valeur pour la période de référence achevée peut être demandée à diverses reprises le jour de référence et qu'à cette occasion l'on peut observer des variations de réponse individuelles, mène ici, tout comme les modèles d'enquêtes répétées associés à un jour de référence, au concept d'une variabilité de réponses individuelle (unité) et totale (collectif). Cette représentation théorique de l'enquête n'est toutefois pas encore praticable. Elle doit une fois de plus être adaptée pour pouvoir calculer à partir de réponses multiples à divers moments de l'enquête une bonne approche de la variance de réponses de la période de référence.

Dans la période $[\tau_1, \tau_2]$ nous voulons connaître des réalisations de l'indice (valeurs de la variable) des unités i d'un mouvement (événements). t_0 , jour de l'enquête de l'enquête principale (moment de l'enquête), se trouve de manière appropriée à la fin de $[\tau_1, \tau_2]$ et est généralement identique à τ_2. Pour estimer la variabilité de réponse individuelle comme variance de réponse, il faut, tout comme pour les états (stocks), demander au moins deux valeurs par unité. Pour les enquêtes répétées nécessaires à cet effet, il faut fixer - dans le temps après t_0 , de nouveaux moments d'enquête t_1, t_2, t_3, etc. Le schéma du modèle flux (mouvements) le montre de manière claire.

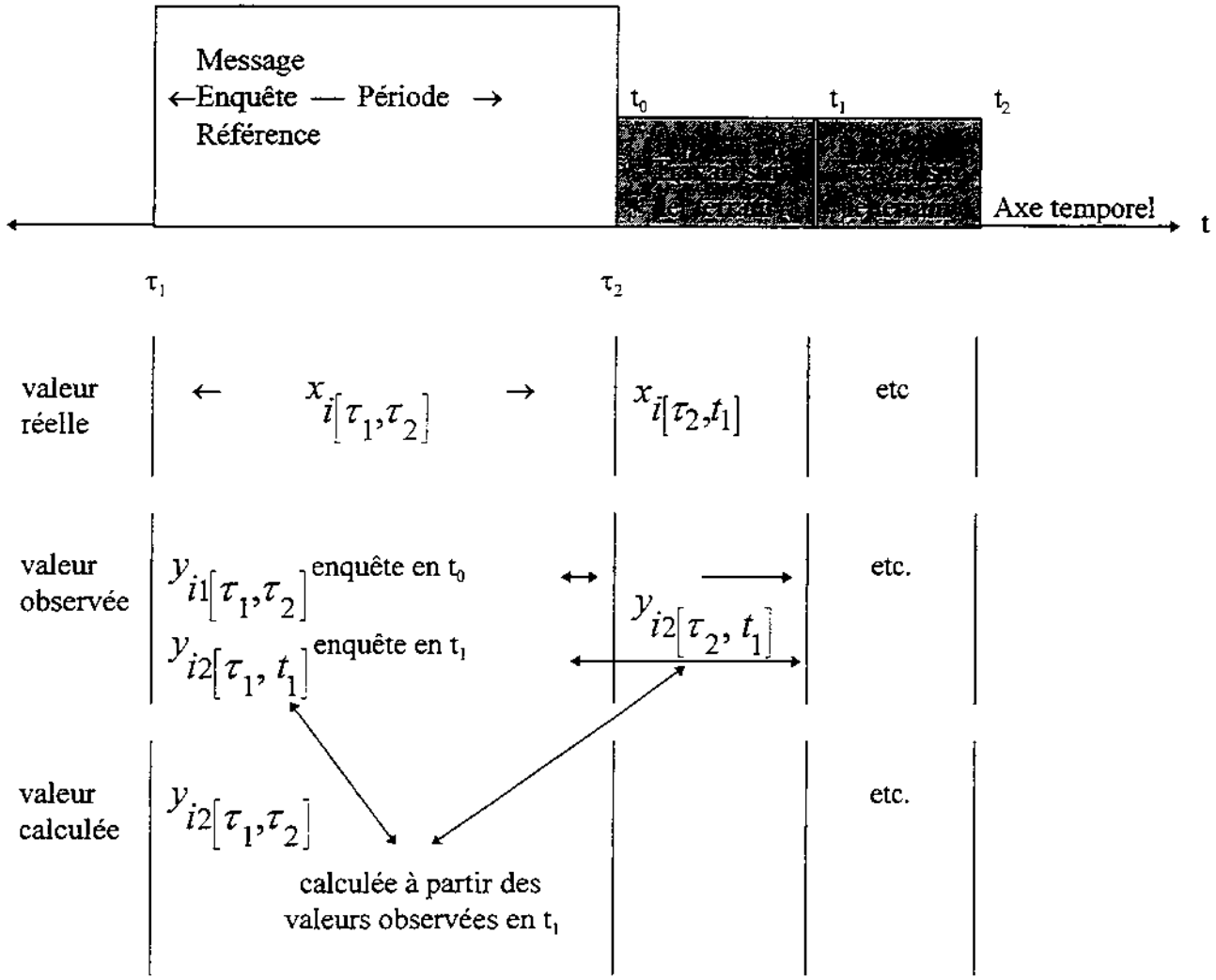


Schéma du modèle "Mouvements" (flux)

Au moment t_1 l'on demande les événements et les réalisations dudit indice de la période $[\tau_1, \tau_2]$. Au moment t_1 l'on peut demander les événements de la période $[\tau_1, t_1]$ et $[t_0 = \tau_2, t_1]$ ou aussi - mais ce n'est qu'un cas de figure - $[\tau_1, \tau_2]$ et $[t_0 = \tau_2, t_1]$. De même pour les moments t_2, t_3, etc. Pour les valeurs périodiques liées à l'unité i, l'on utilise les dénominations suivantes :

$x_i[\tau_1, \tau_2]$	valeur réelle de la variable à $[\tau_1, \tau_2]$;); ne peut généralement pas être établie, sauf en cas d'enregistrement (pas de questionnement). Ici $t_0 = \tau_2$.
$y_{i1}[\tau_1, \tau_2]$	valeur déclarée au moment $t_0 = \tau_2$ pour $[\tau_1, \tau_2]$; enquête principale
$x_i[\tau_2, t_1]$	valeur réelle pour $[\tau_2, t_1]$
$y_{i2}[\tau_1, t_1], y_{i2}[\tau_2, t_1]$	valeurs déclarées à t_1 en $t_0 = \tau_2$ pour $[\tau_1, t_1]$ et $[t_0 = \tau_2, t_1]$, où $t_1 > t_0$; première enquête répétée
$y_{i2}[\tau_1, \tau_2]$	"valeur déclarée calculée" pour $[\tau_1, \tau_2]$ provenant de la première enquête répétée dans l'hypothèse d'une ventilation de l'erreur comme formulée plus loin
$y_{i3}[\tau_1, t_2], y_{i3}[\tau_2, t_2]$	valeurs déclarées à t_2 en $t_0 = \tau_2$ pour $[\tau_1, t_2]$ et $[t_0 = \tau_2, t_2]$, où $t_2 > t_0$; deuxième enquête répétée
$y_{i3}[\tau_1, \tau_2]$	idem pour t_2 comme dans $y_{i2}[\tau_1, \tau_2]$ et t_1
$\varepsilon_{i1}[\tau_1, \tau_2]$	erreur de déclaration dans l'enquête principale pour la période $[\tau_1, \tau_2]$, où $t_0 = \tau_2, \varepsilon_{i1}[\tau_1, \tau_2] = y_{i1}[\tau_1, \tau_2] - x_i[\tau_1, \tau_2]$
$\varepsilon_{i2}[\tau_1, t_1], \varepsilon_{i2}[\tau_2, t_1]$	erreur de déclaration dans la première enquête répétée pour $[\tau_1, t_1], [\tau_2, t_1]$
$\varepsilon_{i2}[\tau_1, \tau_2]$	"erreur de déclaration calculée" de la première enquête répétée
$\tilde{\varepsilon}_{i2}[\tau_2, t_1]$	symbole auxiliaire pour la ventilation de l'erreur, voir l'hypothèse dans le cas $\tilde{\varepsilon}_{i2}$
$\varepsilon_{i3}[\tau_1, t_2], \varepsilon_{i3}[\tau_2, t_2]$	$\varepsilon_{i3}[\tau_1, \tau_2]$ idem pour la deuxième enquête répétée

Des grandeurs provenant d'enquêtes répétées ultérieures sont calculées de manière analogue.

Les valeurs $y_{i1}[\tau_1, \tau_2], y_{i2}[\tau_1, \tau_2], y_{i3}[\tau_1, \tau_2]$ etc. sont généralement différentes. Dans la suite de notre exposé nous partons de valeurs examinées à deux moments de l'enquête t_0, t_1, parce que dans la pratique plus de deux enquêtes répétées sont exclues.

L'on peut former de façon analogue des formules pour plus de deux valeurs examinées. Les rapports suivants valent pour la valeur de l'enquête principale :

$$y_{i1}[\tau_1,\tau_2] := x_i[\tau_1,\tau_2] + \varepsilon_{i1}[\tau_1,\tau_2]$$

la valeur de l'enquête répétée :

$$\begin{aligned} y_{i2}[\tau_1,\tau_2] &= y_{i2}[\tau_1,t_1] - y_{i2}[\tau_2,t_1] \\ &= x_i[\tau_1,\tau_2] + \varepsilon_{i2}[\tau_1,\tau_2] + x_i[\tau_2,t_1] + \varepsilon_{i2}[\tau_2,t_1] - (x_i[\tau_2,t_1] + \tilde{\varepsilon}_{i2}[\tau_2,t_1]) \\ &= x_i[\tau_1,\tau_2] + \varepsilon_{i2}[\tau_1,\tau_2] \end{aligned}$$

la valeur moyenne :

$$\bar{y}_i[\tau_1,\tau_2] = \frac{1}{2}[y_{i1}[\tau_1,\tau_2] + y_{i2}[\tau_1,\tau_2]] = x_i[\tau_1,\tau_2] + \frac{1}{2}(\varepsilon_{i1}[\tau_1,\tau_2] + \varepsilon_{i2}[\tau_1,\tau_2]) \ .$$

Pour cette transformation (une sorte d'"ajustement") l'on a pris pour hypothèse que $\tilde{\varepsilon}_{i2}[\tau_2,t_1] \approx \varepsilon_{i2}[\tau_2,t_1]$. L'on suppose donc sur le plan du contenu que l'erreur pour la période (partielle) $[\tau_2,t_1]$ dans les deux enquêtes (pour les valeurs $y_{i2}[\tau_1,t_1], y_{i2}[\tau_2,t_1]$) est pratiquement identique.

III. Variance des réponses : estimation de la valeur de caractéristiques constantes dans le temps

On effectue des enquêtes répétées afin d'obtenir deux valeurs observées d'une caractéristique de chaque unité d'enquête *i*; à l'aide desquelles l'on pourra estimer une variance des déclarations ou des réponses. Si la valeur de la caractéristique de toutes les unités est constante ou à tout le moins constante pour une période assez longue, il n'est pas nécessaire de transposer ou d'ajuster la valeur examinée à une jour fixé uniformément. A cette fin existent le modèle (b1) et - légèrement modifié - le modèle pour flux. Sans ajustement on dispose de y_{i1}, y_{i2}, qui agissent directement sur $\sigma^2_{R_i}$, formule d'estimation de la variance simplifiée réduite au cas de deux résultats de dénombrement. L'estimateur de la variance de réponses individuelle de l'unité *i* est indiqué dans la limitation $k = 2$.

$$\sigma^2_{R_i} \approx s^2_{R_i} = \frac{1}{2-1}\left[(y_{i1} - \frac{y_{i1}+y_{i2}}{2})^2 + (y_{i2} - \frac{y_{i1}+y_{i2}}{2})^2\right] = \frac{1}{2}(y_{i1} - y_{i2})^2 = \frac{d_i^2}{2}$$

La variance des réponses σ^2_R est estimée comme moyenne des variances des réponses individuelles

$$\sigma^2_R \approx s^2_R = \frac{1}{N}\sum_{i=1}^{N} s^2_{R_i}$$

N représente ici le nombre d'unités dans l'univers. La variabilité des réponses V_R ou aussi variance d'échantillonnage des réponses est estimée par la variance d'échantillonnage des réponses $\hat{V}_R$:

$$V_R = \frac{\sigma^2_R}{2N} \approx \hat{V}_R = \frac{s^2_R}{2N}$$

Lorsque l'enquête principale et l'enquête répétée sont des sondages, les formules de sondage correspondantes sont valables.

A côté de beaucoup d'autres exemples de caractéristiques constantes, tels que la date de naissance, la situation familiale, la statut juridique de l'exploitation, le statut professionnel, etc., l'on peut dans les statistiques agricoles considérer aussi comme valeur de caractéristique constante : l'équipement en machines d'une production durant une certaine période, ou la surface cultivée des exploitations agricoles pendant une période particulière de végétation.

Pour ce qui est de ce dernier point, l'on a utilisé en Belgique, pendant l'été 1985, lors du Recensement agricole et horticole, un modèle d'enquête répétée pour déterminer la variance des réponses dans les déclarations relatives aux surfaces cultivées qu'avaient effectuées les exploitations.

Le modèle d'enquête répétée comportait une enquête principale et une enquête répétée. L'enquête principale était une "enquête exhaustive", le recensement agricole et horticole du 15 mai 1985 (t_0); l'enquête répétée était ce qu'on appelle une "enquête-pilote ex ante" qui a eu lieu le 4 mai 1985 (t_{-1}). Les questionnaires se trouvent en annexe.

Pour toutes les superficies cultivées de cultures principales, relevées dans l'enquête principale (céréales, plantes industrielles, total général de la superficie agricole utilisée), l'on a appliqué par approximation, selon le même système exécutoire, le modèle répété, limité, pour des raisons financières et organisationnelles, a deux des sept catégories d'exploitation, c'est-à-dire :

Catégorie 1 : les personnes dont la profession principale est l'agriculture ou l'élevage;

Catégorie 5 : les personnes, exploitations ou organismes qui produisent en vue de la vente des produits végétaux et offrent en même temps à d'autres exploitations des services agricoles.

Dans toutes les catégories l'enquête principale a été effectuée comme une enquête ou recensement exhaustif; l'enquête-pilote par contre était un sondage (choix aléatoire de quelque 4 % des exploitations) dans la catégorie 1 et une enquête ou recensement exhaustif dans la catégorie 5. Cette différenciation était nécessaire au motif que la catégorie 1 comptait quelque 64000 exploitations et la catégorie 5 seulement quelque 450 exploitations. La sélection s'est faite d'après les exploitations du Recensement agricole et horticole belge du 15 mai 1983, complétées par les nouvelles exploitations au 15 mai 1985.

L'enquête répétée s'est faite sur une base volontaire. L'on a envoyé par la poste les mêmes questionnaires que ceux de l'enquête principale (même mode d'exécution, sans effet recenseur). Le quota des non-réponses était de quelque 11,5 %, compte tenu des exploitations qui avaient fermé entre le 15 mai 1983 et le 15 mai 1985; si l'on en croit l'expérience de la statistique officielle de l'époque, il s'agit dans le cas d'une enquête facultative comme celle-ci d'un fort bon résultat qui ne déforme pas fondamentalement les calculs provenant de ces données. La surcharge que représentait d'un point de vue organisationnel deux enquêtes dans un laps de temps assez court et le quota relativement bas des non-réponses ont fait que l'on a sciemment renoncé à analyser la structure des exploitations qui n'avaient pas répondu et à analyser les données manquantes correspondantes que l'on avait complétées.

Afin de limiter les coûts et de ne pas trop nuire aux autres travaux de l'Institut, l'on n'a poursuivi l'analyse que pour les cultures principales les plus importantes, à savoir dans la :

Catégorie 1 : les surfaces cultivées de froment d'hiver, d'orge d'hiver, de betteraves et de plantes industrielles, et la superficie agricole utilisée;

Catégorie 5 : la surface cultivée de froment d'hiver et la superficie agricole utilisée.

Dans cette première partie, nous donnons quelques exemples de calculs effectués d'après des déclarations de la catégorie 5. Le coût total du traitement est limité vu le nombre peu élevé d'exploitations dans cette catégorie. Dans l'enquête-pilote toutes les exploitations ont été soumises à une enquête appro-fondie. De cette façon l'on a pu éliminer l'erreur d'échantillonnage aléatoire complémentaire propre au choix d'un échantillon d'exploitations. De la sorte, l'estimation de la variance des réponses moyenne et de la variabilité des réponses n'est pas en plus encombrée par des effets de sondage, mais représentée seulement comme l'estimation des déclarations répétées pour l'ensemble ($k = 2$). Le modèle de la catégorie 5 est donc un modèle parfait d'erreur de réponse sans choix aléatoire d'échantillonnage et sans stratification.

Conformément au programme de tableaux prévu le traitement et l'évaluation ont été effectués d'après quatre classes de grandeur de la superficie agricole utilisée; cf. tableau arithmétique 1a/b et "exemple du mode de calcul pour la catégorie 5". A titre d'exemple (de l'ensemble des résultats) l'on a établi ici les résultats relatifs au froment d'hiver.

Comme nous l'avons signalé plus haut, les valeurs de caractéristiques que nous avons prises étaient constantes pour la période de végétation printemps/été 1985. Cette constance dans le temps des valeurs de caractéristiques à examiner permet d'utiliser un modèle sans ajustement. Ceci crée une situation favorable pour l'estimation des variances des réponses du fait qu'entre les jours de collecte il n'y a pas de changements dans les valeurs de caractéristiques. De la sorte se dissipent les effets qui auraient pu résulter de l'application d'un ajustement. Dans la classe de grandeur 0,01 ha - < 10 ha l'on a estimé pour la surface cultivée froment d'hiver une variabilité des réponses relative (racine) de 1,5 %. Dans les classes de grandeur suivantes (2 à 4) (pour la définition des classes de grandeur, voir le tableau arithmétique 1b) l'on a trouvé comme valeurs correspondantes 2,3 %, 3,6 % et 0 %. La variabilité des réponses en ce qui concerne la surface cultivée froment d'hiver de toutes les exploitations de la catégorie 5 a été estimée à 1,5 %, la valeur correspondante pour la superficie agricole utilisée à 4,1 %.

La variance des réponses peut aussi être interprétée comme une variance d'une répartition de déclaration "moyenne" composée de la répartition de déclaration individuelle. Les résultats de ces calculs pour la catégorie 5 (activité agricole secondaire) se retrouvent dans les tableaux 8a et 8b.

Bien que l'on puisse supposer, particulièrement en ce qui concerne les surfaces cultivées, que la variabilité des réponses dans le cas de valeurs de caractéristiques constantes dans le temps soit relativement peu importante, les valeurs de la catégorie 5 présentent à ce propos une particularité que l'on peut attribuer au caractère propre de cette catégorie (activité agricole secondaire) à effectif relativement réduit.

Des résultats correspondants de la catégorie 1 (l'agriculture est l'activité principale - voir partie 2, tableau 3) ont donné en ce qui concerne la variabilité des réponses des valeurs intrinsèquement moindres : 0,5 % pour le froment d'hiver et 0,2 % pour la superficie agricole utilisée.
Ce dernier ordre de grandeur correspond mieux aux représentations qu'on se fait de la variabilité des réponses dans le cas de valeurs de caractéristiques constantes dans le temps ayant à la fois un grand effet de rappel. Toutefois les résultats de la catégorie 5 démontrent que des caractéristiques constantes dans le temps et dont l'effet de rappel est plus faible ne doivent absolument pas aller de pair avec une variabilité des réponses moindre.

Catégorie 5 - exemple de calcul

Classe de grandeur 1 : 0,01 ha - < 10 ha de superficie agricole utilisée (SAU)
Province 2 (Brabant), commune B - 24008 Bekkevoort
L'exploitation (i) n° 0467 dans cette commune avait effectué, lors du Recensement agricole et horticole (15 mai 1985) et lors de l'enquête-pilote (4 mai 1985), les déclarations suivantes - en ares - en ce qui concerne le code 014 (superficie froment d'hiver) : 4 mai 1985 : $y_{i2} = 340$ a ; 15 mai 1985 : $y_{i1} = 351$ a.

Les calculs ont donné les résultats suivants :

Déviation des valeurs déclarées $d_i = y_{i1} - y_{i2} = 351 - 340 = 11$ a

Variance des réponses individuelle : $s^2_{R,1} = \frac{d_i^2}{2} = \frac{11^2}{2} = 60{,}5$

Il va de soi que l'on ne peut que calculer les déviations des unités individuelles i pour lesquelles l'on dispose à chaque fois de valeurs déclarées (y_{i1}, y_{i2}) relatives à l'enquête principale et à l'enquête répétée (enquête-pilote); en bref appelé "effectif".

Comme dans l'enquête-pilote l'on a interrogé toutes les exploitations de cette catégorie - composé de l'ancien fichier majoré de toutes les nouvelles exploitations créées entre le 15 mai 1983 et le 15 mai 1985 - il n'est pas nécessaire lors de l'estimation de la variance des réponses et de la variabilité des réponses de faire une distinction entre l'ancien fichier et les nouvelles exploitations (au contraire des estimations concernant la catégorie 1); par exploitations de l'ancien fichier l'on entend celles qui étaient reprises dans le recensement du 15 mai 1983 et qui ont continué à fonctionner durant la période allant du 15 mai 1983 au 15 mai 1985.

Ceci simplifie substantiellement l'estimation.

Commentaire :

Catégorie 5 : Exploitations (personnes, institutions, exploitations) qui, à titre principal ou secondaire, effectuent des travaux agricoles et horticoles pour le compte d'agriculteurs ou d'horticulteurs, ou mettent à la disposition de ces derniers des machines ou installations agricoles ou horticoles, et produisent elles-mêmes en même temps des produits végétaux ou animaux destinés à la vente.

Ancien fichier : comprend les exploitations qui existaient au moment des recensements agricole et horticole des 15 mai 1983 et 15 mai 1985.

Nouvelles exploitations : comprend celles qui n'existaient pas encore à l'époque du recensement agricole et horticole du 15 mai 1983 et ont été créées entre ce dernier et l'enquête-pilote du 4 mai 1985 et le recensement du 15 mai 1985.

Renvoi aux formules :

Comme la catégorie 5 a été interrogée de manière exhaustive, l'on a pu simplifier les autres formules d'estimation pour répartir selon les quatre classes de grandeur la superficie agricole utilisée L = 4 (estimation ponctuelle et intervalle d'estimation) :

Variance des réponses

$$S_R^2 = \frac{1}{N}\sum_h N_h \frac{\sum_i S_{R,h}^2}{N_h} = \frac{1}{N}\sum_h\sum_i S_{R,h}^2 \qquad (i = 1, 2, ..., N_h)$$

$$(h = 1, 2, ..., L\ ;\ L = 4)$$

Variabilité des réponses ou variance d'échantillonnage des réponses

$$V_R \approx \frac{1}{N^2}\sum_h N_h^2 \cdot \frac{\frac{1}{N_h}\sum_i S_{R,h}^2}{2N_h} = \frac{1}{2N}\cdot\frac{1}{N}\sum_h\sum_i S_{R,h}^2 = \frac{S_R^2}{2N}$$

(Pour les formules relatives à la stratification, voir W.G. COCHRAN, Stichprobenverfahren (version allemande), Berlin - New York, 1972, pp. 111-118, ou l'édition anglaise intitulée, Sampling Techniques, 3rd Ed., Wiley, New York and London, 1977, pp. 89-96).

Variance des réponses et variabilité des réponses de la 1ère classe de grandeur 0,01 ha - < 10 ha S.A.U. (ancien fichier + nouvelles exploitations créées; c'est-à-dire toutes les exploitations de la classe de grandeur 1, voir tableaux de calcul 1a et 1b).

Variance des réponses (estimation pour l'ensemble des exploitations à partir de k= 2 valeurs déclarées pour chaque exploitation)

$$S_{R1}^2 = \frac{1}{N_{1(85)}}\left(\sum_i^{N_{1(85)}} S_{R,1}^2\right) = \frac{1}{266}\ 55089{,}5 = 207{,}103$$

Variabilité des réponses, variance de l'échantillonnage des réponses (estimation de la variance moyenne à partir des valeurs moyennes provenant de k = 2 valeurs déclarées)

$$V_{R1} = \frac{S_{R1}^2}{2N_{1(85)}} = \frac{207{,}103}{2 \cdot 266} = 0{,}389$$

(Valeur relative après extraction de la racine)

$$V_{R1}^{1/2}(\%) \approx \frac{S_{R1}}{\sqrt{2N_{1(85)}}} \Big/ \frac{Y_{21} + Y_{11}}{2N_{1(85)}}(\%) = \frac{0{,}624}{41{,}553} = 1{,}5\ (\%)$$

Calculs pour l'ensemble des exploitations de la catégorie 5; voir tableau de calcul 1b :

Variance des réponses

$$S_R^2 = \frac{1}{N_{(85)}} \sum_{h}^{4} \left(\sum_{i}^{N_{h(85)}} S_{R,h}^2 \right) = \frac{1}{378} 1790761{,}5 = 4737{,}464$$

Variabilité des réponses, variance de l'échantillonnage des réponses (estimation de la variance moyenne à partir des valeurs moyennes provenant de $k = 2$ valeurs déclarées)

$$V_R \approx \frac{S_R^2}{2N_{(85)}} = \frac{4737{,}464}{2 \cdot 378} = 6{,}266$$

(Valeur relative après extraction de la racine)

$$V_R^{1/2}(\%) \approx \frac{S_R}{\sqrt{2N_{(85)}}} / \frac{Y_2 + Y_1}{2N_{(85)}}(\%) = \frac{2{,}503}{162{,}414} \approx 1{,}5(\%)$$

L'attribution des exploitations à la classe de grandeur de la superficie agricole utilisée (S.A.U.) s'est faite sur la base des valeurs déclarées (S.A.U.) de l'enquête principale (15 mai 1985).

Tableau 1

Recensement agricole et horticole belge - mai 1985
Définition de la variance des réponses et de la variabilité des réponses sans ajustement
Catégorie 5 - Enquête exhaustive; caractéristique : code 014 superficie en froment d'hiver - Calculs
Table de calcul a)
Classe de grandeur O,01 ha - < 10 ha ou 1 a - < 10,00 a - Superficie agricole utilisée (SAU)
- Extrait -

N° d'ordre des exploi-tations (i)	Code de l'exploitation i correspondant au n° de la commune N° de l'exploitation (i)	Valeurs déclarées en a		$d_i = y_{i1} - y_{i2}$	Variance des réponses individuelles $s^2_{R_i1} = \frac{d_i^2}{2}$
		4 mai y_{i2}	15 mai y_{i1}		
		- Ancien fichier (A.F.) -			
	1. Province d'Anvers				
1	11009 Brecht / 0360	0	0	0	0
2	11053 Wuustwezel / 0446	0	0	0	0
.					
.	2. Province de Brabant				
.	.				
.	.				
.	.				
26	24008 Bekkevoort/0467	340	351	11	60,5
.	.				
.	.				
32	24045 Huldenberg/0396	240	240	0	0
.	.				
.	.				
40	25048 Jodoigne/0131	90	50	-40	800,0
.	.				
.	.				
.	9. Province de Namur				
256	92006 Assesse/0096	1000	935	-65	2112,5
.	.				
.	.				
259	92035 Eghezée/0340	0	0	0	0
	Au total : N_{AF1} = 259 exploitations	$\sum y_{i2}$ = 10358	$\sum y_{i1}$ = 10491	$\sum d_i$ = 153	$\sum s^2_{R_i1}$ = 55089,5
		- Nouvelles exploitations créées (N.F.) -			
	2. Province de Brabant				
1	24014 Boortmeerbeek/0217	0	0	0	0
.	.				
.	.				
	6. Province de Liège				
6	64063 Remicourt/0120	229	229	0	0
	7. Province du Limbourg				
7	71022 Hasselt/0535	150	150	0	0
	Au total N_{NE1} = 7 exploitations	$\sum y_{i2} = 629$	$\sum y_{i1} = 629$	$\sum d_i = 0$	$\sum s^2_{R_i1} = 0$
.	total : $N_1 = N_{AF1} + N_{NE1}$	$Y_{21} = \sum y_{i2}$	$Y_{11} = \sum y_{i1}$	$\sum d_i$	$\sum s^2_{R_i1}$
.	= 266 exploitations $N_{1(85)}$	= 10987	= 11120	= 153	= 55089,5

Abréviations : ha = hectare, a = are

Tableau 1
Table de calcul b)

Valeurs auxiliaires pour le calcul de la variance des réponses S^2_{Rh} dans la classe de grandeur h et de la variance des réponses S^2_R						Calcul des variances des réponses S^2_{Rh} et S^2_R et des variabilités des réponses V_{Rh} en V_R (variance d'échantillonnage des réponses)			
Classe de grandeur de la supperficie argicole utilisée en ha h	Nombre d'exploitations (effectif) $N_{h(85)}$	Valeurs totales des déclarations (effectif) en a 4 mai Y_{2h}	15 mai Y_{1h}	$\sum_i d_{ih}$	$\sum_i s^2_{R_i h}$	Moyenne du total $\frac{Y_{2h}+Y_{1h}}{2N_{h(85)}}$ in a	Variance de réponses $S^2_{Rh} = \frac{1}{N_{h(85)}}\sum_i S^2_{R_i h}$	Variabilité des réponses V_{Rh} $\frac{S^2_{Rh}}{2N_{h(85)}}$	Racine carée %
0,01 - < 10	266	10987	11120	153	55089,5	41,553	207,103	0,389	1,5
10 - < 30	87	26003	25621	- 382	711749	296,690	8181,023	47,017	2,3
30 - < 100	24	19355	20699	1344	1023923,0	834,458	42663,458	888,822	3,6
100 - < +	1	4500	4500	0	0	4500	0	0	0,0
Total	$N_{(85)}$ = 378 nombre d'exploitations en catégorie 5	$Y_2 = \sum Y_{2h}$ = 60 845	$Y_1 = \sum Y_{1h}$ = 61 940 Total superficie froment (effectif) en a	$\sum_h \sum_i d_{ih}$ = 1315	$\sum_h \sum_i s^2_{R_{ih}}$ =1790761,5	$\frac{Y_2+Y_1}{2N_{(85)}}$ = 162,414 in a	S^2_R = 4737,464	$\frac{S^2_R}{2N_{(85)}}$ = 6,266	1,5

Ici nous avons $Y_{2h} = \sum_{i=1}^{N_{h(85)}} y_{i2}$; $Y_{1h} \sum_{i=1}^{N_{h(85)}} y_{i1}$:

A remarquer qu'il n'y a pas eu d'estimations séparées pour l'ancien fichier des exploitations et pour les nouvelles exploitations créées : l'enquête principale et l'enquête répétée (enquête-pilote) ont été effectuées comme des recensements exhaustifs ($n = N$) ; donc sans effet d'échantillonnage dans le sens strict de la théorie des sondages !

IV. Variance des réponses : estimation de la valeur de caractéristiques non-constantes dans le temps - méthodes d'ajustement dans le temps

Dans le cas de caractéristiques d'états (stocks) qui, en règle générale, connaissent à court terme des fluctuations - comme par exemple le nombre de travailleurs dans une exploitation, le cheptel dans les exploitations agricoles ou encore l'importance des stocks - l'on n'a pas à faire à des valeurs de caractéristiques constantes entre deux jours de sondage. L'ajustement des valeurs recueillies lors de plusieurs jours de sondage en un jour de sondage unique est une condition indispensable pour l'estimation des variances des réponses. Sur cette base un développement approprié et en même temps réalisable de l'ajustement détermine la validité des résultats obtenus.

L'ajustement lors d'enquêtes répétées où les personnes interrogées remplissent elles-mêmes les feuilles (sans l'appui du recenseur) doit de manière formelle être représenté comme suit : dans un univers avec N unités dont les caractéristiques ayant un intérêt statistique sont variables dans le temps il faut enregistrer les valeurs de caractéristiques au jour de sondage t_O. La forme de l'enquête, soit recensement exhaustif, soit sondage, est définie dans le système de travail. Pour estimer la variance des réponses il faut au moins deux valeurs de la même caractéristique, recueillie le même jour de sondage t_O. Pour cette raison il faut qu'au moins deux enquêtes échelonnées dans le temps se suivent. La méthode de conversion est intitulée ajustement conformément à la définition suivante :

Déf. : L'ajustement est une méthode par laquelle s'effectue la conversion d'une valeur de caractéristique y^*_{i2} appartenant au jour de sondage t_1 en une valeur y_{i2} appartenant au jour de sondage t_O. La conversion peut s'effectuer de façon rigoureuse (si l'on dispose des valeurs réelles des modifications de la valeur de caractéristique intervenues entre temps) ou par estimation (auquel cas les conditions du système de travail utilisé doivent accessoirement être prises en considération).

Pour effectuer les enquêtes en pratique, l'on a développé et testé le modèle b2 (voir première partie, section II C 1) au moyen d'un ajustement spécial adapté (STRECKER, WIEGERT, 1984).

Dans ce modèle b2 pragmatique l'ajustement peut de fait se faire de quatre façons différentes dont l'une, si exacte soit-elle, est en pratique peu réalisable et les trois autres représentent des méthodes d'évaluation. La condition pour l'ajustement par estimation est que les modifications des valeurs de caractéristiques entre les jours de sondage ne soient pas purement l'effet du hasard mais aient des composants systématiques qui, comme il est admis communément dans le cas de modèles de séries chronologiques, soient couverts par des influences non-dominantes dues au hasard.
Pour appliquer une méthode d'estimation il faut vérifier si les données empiriques justifient l'application d'une telle prémisse. Si tel est le cas, l'on peut fixer entre les jours de sondage t_0, t_1 l'estimation d'un rapport de variation moyen de la caractéristique, voire des valeurs de caractéristique.
Ci-après, sont présentés les quatre ajustements les plus importants :

(1) L'on relève avec précision pour chaque unité i les modifications individuelles entre la valeur y_{i2} de l'effectif au jour de sondage t_0 et la valeur y^*_{i2} de l'effectif au jour de sondage t_1.
Il s'ensuit : $y_{i2} = y^*_{i2} + \Delta_i$; Δ_i étant la modification entre les valeurs réelles x_i, x^*_{i2}. Les valeurs y^*_{i2} et Δ_i permettent ainsi d'ajuster au jour de sondage t_0 une seconde valeur observée. Cette façon d'ajuster est utilisable, à tout le moins en théorie, tant pour les unités d'un recensement exhaustif que pour les unités d'un modèle de sondage à un ou plusieurs degrés. En pratique, la détermination précise des modifications dans le temps des valeurs réelles se heurtera pour chaque unité - lors d'une troisième enquête supplémentaire - à d'importantes difficultés au motif que ces enquêtes concernant toutes les unités sont en fait irréalisables à des moments rapprochés. Cette forme d'ajustement précis doit pour cette raison être remplacée en règle générale par des estimations.

(2) L'ajustement est effectué à l'aide de méthodes basées sur des séries chronologiques. Cette sorte d'ajustement n'est avancée que pour la forme; elle implique de telles hypothèses extrêmes que sur base des données disponibles il est quasiment impossible de les mettre en pratique. Pour pouvoir analyser dans le temps les modifications individuelles de la valeur de caractéristique de chaque unité i, il faut disposer d'un grand nombre de valeurs d'unité "réelles" de ce type dans le passé, auxquelles il faut appliquer une méthode d'extrapolation appropriée pour assurer la continuité des valeurs du moment t_0 au moment t_1 sur la partie actuelle de la série chronologique. Dans la pratique toutefois l'on ne disposera pas d'une aussi vaste information concernant les valeurs individuelles et leur analyse temporelle.

Cette forme d'ajustement n'est qu'une hypothèse théorique comme nous l'avons déjà démontré dans la première partie, section II A. Si l'on connaît à tout le moins quelques modifications réelles de valeurs de caractéristiques individuelles du passé pour quelques unités de l'enquête, par exemple à intervalles égaux, l'on peut estimer une modification journalière moyenne de la caractéristique et l'utiliser pour l'ajustement des valeurs y^*_{i2} à y_{i2} de t_1 à t_O. Cette forme d'ajustement suppose que l'estimation du passé est transmissible à la situation actuelle et que les rapports entre les modifications des variables individuelles dans le temps n'ont pas changé de manière fondamentale. Même alors cette méthode n'autorise qu'une approche vraiment rudimentaire.

(3) En ce qui concerne les unités de l'enquête l'on connaît une autre caractéristique auxiliaire, comme par exemple leurs valeurs au moment t_O, et l'on sait en plus que les valeurs de ces caractéristiques auxiliaires sont en étroite corrélation avec le schéma modificateur dans le temps de la caractéristique de l'enquête (corrélation moyenne à élevée). Afin de rendre ce savoir utilisable, l'on forme si possible des classes de grandeur homogènes d'après la caractéristique auxiliaire et l'on répartit les unités dans les classes de grandeur. De chaque classe de grandeur l'on tire proportionnellement à la valeur de la caractéristique auxiliaire un (sous)-échantillon dont la taille est définie par rapport à la classe occupée. Dans les unités choisies de la sorte une enquête de contrôle permet d'établir les modifications dans le temps de la valeur réelle. A l'intérieur de chaque classe l'on suppose par approximation qu'il y a un lien linéaire entre les modifications $\Delta_i = x_i - x^*_{i2}$ et les valeurs de la caractéristique auxiliaire. L'on effectue cette estimation pour chaque classe. De la sorte l'on obtient pour toutes les unités de la classe, des estimations Δ_i pour les variations réelles individuelles que l'on utilise à des fins de correction :

$$y_{i2} \approx y^*_{i2} + \hat{\Delta}_i$$

L'information auxiliaire peut dans le cas de grandeurs d'un effectif, par exemple, être la variable d'enquête elle-même, voire leur valeur lors d'enquêtes précédentes. L'on peut alors dans des classes de grandeur homogènes, constituées judicieusement et pas trop vastes, sélectionner des unités et les soumettre à l'enquête de contrôle. La troisième enquête complémentaire nécessaire (enquête de contrôle) ne sera pas trop importante pour des raisons de temps, de coût et d'organisation.

Les trois cas que l'on vient de décrire traitent du problème de l'ajustement sans prendre en considération les limitations qu'impose en pratique l'enquête. Ils offrent chacun une vision théorique possible.

La prise en considération des exigences particulières que pose la faisabilité d'enquêtes répétées mène au modèle C (comme déjà décrit dans la partie I section II).

Dans la quatrième sorte d'ajustement qui suit, seront traités de manière générale les problèmes pragmatiques.

(4) Le modèle (c) comprend une enquête principale (au moment t_0), une enquête répétée ex ante (au moment t_{-1}) et une autre enquête répétée à l'issue du travail sur le terrain de l'enquête principale. Il n'est pratiquement pas possible d'observer directement les modifications intervenues dans les valeurs de caractéristiques de chaque unité i : $\Delta_i = y_{i2} - y_{i2}^{\bullet}$ car il faudrait pour cela effectuer, de manière simultanée et dans un laps de temps relativement court $[t_{-1}, t_0]$, trois enquêtes qui inévitablement s'influenceraient l'une l'autre. Ce n'est que dans le cours de l'enquête répétée que l'on peut observer les Δ_i ($i = 1,..,N$ dans l'enquête exhaustive ; $i = 1,..,n$ dans le sondage), à savoir comme des modifications dans le temps après l'enquête principale. Pour ajuster t_{-1} à t_0 on a toutefois besoin des valeurs Δ_i des intervalles ex ante $[t_{-1}, t_0]$. L'on bute ici sur un obstacle fondamental, inhérent à l'utilisation de la méthode des recensements répétés et ne pouvant être surmonté qu'à l'aide de nouvelles hypothèses basées sur la réalité des données. Si l'on ne parvient guère à former des hypothèses plausibles faisant apparaître comme rationnelle et réalisable une estimation de Δ_i dans un intervalle ex ante, l'application du modèle (c) pour déterminer la variabilité des réponses n'est pas possible lorsqu'il s'agit de valeurs de caractéristiques variant dans le temps.

Il est malaisé de se représenter une autre solution pratique. Pour définir concrètement la variance des réponses il faut poser le problème de l'ajustement de manière double, c'est-à-dire comme un problème de "reflet dans le temps" des modifications Δ_i et comme un examen du caractère aléatoire des modifications individuelles Δ_i .

La transmission d'une modification individuelle moyenne de la période de l'enquête répétée dans l'intervalle ex ante au calcul d'une modification journalière moyenne est alors possible lorsqu'il existe des indications que ce "reflet dans le temps" peut être effectué sans commettre de trop grosses erreurs.

Lorsque, par exemple, un composant systématique des variations de valeur dans le temps est présent et reconnaissable, celui-ci peut être estimé au moyen du Δ_i enquêté et transposé dans la période précédente $[t_{-1}, t_0]$. Les valeurs du jour de référence t_{-1} peuvent dès lors être ajustées au jour de référence t_0.

Si dans les changements individuels un tel composant systématique n'est pas reconnaissable immédiatement et si, de plus, de fortes influences aléatoires donnent l'impression d'une important manque de régularité, l'application du système à plusieurs degrés que souvent l'on retrouve dans les sondages peut offrir une aide. Dans le cas, par exemple, d'un sondage à deux degrés les changements individuels au second degré peuvent fortement dépendre du hasard, alors que, par contre, les changements du premier degré agrégés, c'est-à-dire résumés dans une somme, décrivent mieux, grâce aux moyennes, les mécanismes d'une composante temporelle systématique. Il est alors indiqué d'effectuer les ajustements au moyen des estimations du premier degré. Cette forme précitée d'ajustement, à savoir :

- conversion pour le changement vers un niveau supérieur d'agrégation
- estimation (approchante) de la composante systématique du Δ_i
- "reflet" de la tendance au moyen de ces agrégats

a été dans un cas concret testée avec succès pour le modèle (c) (STRECKER, WIEGERT, PEETERS, etc., 1983; STRECKER, WIEGERT, 1984). Les résultats en seront communiqués plus loin.

Pour en terminer avec les explications des différentes formes d'ajustement applicables, il faut remarquer que, à chaque fois, la méthode doit être adaptée, de manière flexible et quand même précise, à la nature des caractéristiques examinées, des données disponibles et du modèle de sondage utilisé. Ce n'est qu'ainsi que, lors de recherches pratiques, l'on peut limiter les problèmes pouvant surgir d'une triple enquête effectuée dans les plus brefs délais.

Il ressort alors que l'ajustement est un outil indispensable quand on utilise les modèles d'enquêtes répétées pour définir la variabilité des réponses.

L'ajustement traité en (4) a été appliqué de manière concrète dans une étude spécialisée d'estimation de la variance des réponses lors d'enquêtes sur le cheptel en Belgique en 1979 (STRECKER, PEETERS, WIEGERT, 1985; de même que STRECKER, WIEGERT, PEETERS, etc., 1983).

Ci-après, nous récapitulons les principales caractéristiques de cette méthode.

Dans l'enquête de contrôle (moment après t_0), le service d'inspection de l'Institut national de statistique a tenu compte, pour ce qui est des communes sondées et de leurs exploitations sélectionnées, des changements individuels réels Δ_{ij} (i la commune = unité primaire ; j l'exploitation = unité secondaire) Ceux-ci présentaient de fortes déviations aléatoires et n'étaient de ce fait pas appropriés à l'ajustement individuel des valeurs y^*_{i2} recueillies ex ante.

Afin d'établir le composant systématique et, en même temps, d'en limiter l'influence aléatoire, l'on a estimé des changements journaliers moyens pour l'unité primaire "commune" sur la base des exploitations sondées m_i parmi l'ensemble des exploitations M_i ($m_i \leq M_i$) dans chaque commune sélectionnée :

$$\hat{\Delta}_i = \frac{M_i}{m_i} \sum_{j=1}^{m_i} \Delta_{ij}$$

et pour celles où l'enquête est incomplète :

$$\hat{\Delta}_i = \sum_{j=1}^{M_i} \Delta_{ij}$$

Ces grandeurs plus stables ont été utilisées pour ajuster les valeurs communes des unités primaires y_{i2}^{*} (valeurs communales de l'enquête répétée ex ante), en tenant compte du nombre de jours entre l'enquête répétée et l'enquête principale. A partir des deux valeurs examinées y_{i1} et y_{i2}, ainsi disponibles, une variance des réponses a été estimée pour chaque unité primaire (commune) et ensuite appliquée à chaque exploitation grâce à une variance des réponses estimée.

La variance des réponses en ce qui concerne les valeurs déclarées du cheptel porcin a été estimée sur la base des trois enquêtes suivantes :

(1) Une enquête principale (Recensement agricole et horticole, le 15 mai 1979 étant le jour d'enquête t_0 = 15 mai 1979) en tant que recensement exhaustif.

(2) Une enquête répétée ex ante : enquête-pilote en remplacement d'un véritable recensement répété, le 4 mai 1979 étant le jour d'enquête t_{-1} = 4 mai 1979, en tant que sondage (sondage à deux degrés : le premier degré étant la commune avec n = 130 communes sélectionnées et une sélection à probabilité P_i (avec remise), et effectivement 88 communes différentes; le second degré étant l'exploitation (j) normalement avec m_i = 8 exploitations). Comme dans une commune l'enquête n'a pu avoir lieu, n est tombé à 129. La sélection des exploitations s'est faite parmi l'ensemble de ceux-ci sans distinction de catégorie.

(3) Une enquête de contrôle comme sondage (plan de sondage comme ci-dessus sous (2)) à effectuer après le travail sur le terrain de l'enquête principale.

Les questionnaires de ces trois enquêtes ont été publiés dans STRECKER, WIEGERT, PEETERS, etc., 1983 et STRECKER, PEETERS, WIEGERT, 1985.

La caractéristique "nombre de porcs" de l'exploitation n'est pas constante dans le temps vu que le nombre d'animaux augmente et baisse à un rythme rapide. C'est pourquoi il a été nécessaire d'effectuer à l'aide des changements quotidiens moyens au niveau des unités primaires - communes (i) - l'ajustement des valeurs enquêtées de t_{-1} = 4 mai 1979 à t_O = 15 mai 1979, la période nécessaire à l'ajustement était de $T = \Delta_{1t} = t_O - t_{-1}$ = 11 jours. Les résultats des enquêtes et des conversions , y compris la définition des variances des réponses, ont été, sur la base de symboles antérieurs, communiqués dans les tableaux 2a/b et dans un exemple mathématique (pour de plus amples informations, voir STRECKER, PEETERS, WIEGERT, 1985; STRECKER, WIEGERT, 1986; BECKMANN, WIEGERT, 1987); dans la publication de 1987 l'ajustement et la définition de l'erreur quadratique moyenne et de la variance des réponses ont été représentés dans un schéma (déviation type et variance des réponses dans l'article "Définition et estimation de la variance des réponses").

Pour mieux comprendre l'exemple mathématique suivant nous reproduisons à nouveau sous forme de schéma le modèle d'erreur pour définir la variance et la variabilité des réponses. L'erreur type moyenne elle-même (Erreur quadratique moyenne ou EQM) peut être calculée - ou bien pour la valeur totale estimée comme ici

$\hat{Y} = \frac{\hat{Y}_2 + Y_1}{2}$ respectivement $\hat{Y}' = \frac{\hat{Y}_2 + \hat{Y}_1}{2}$, ou bien pour la moyenne de l'échantillon

$\bar{y}_{(2)n} = \frac{1}{n}\sum_{i=1}^{n} \frac{y_{i1} + y_{i2}}{2}$ de la sorte :

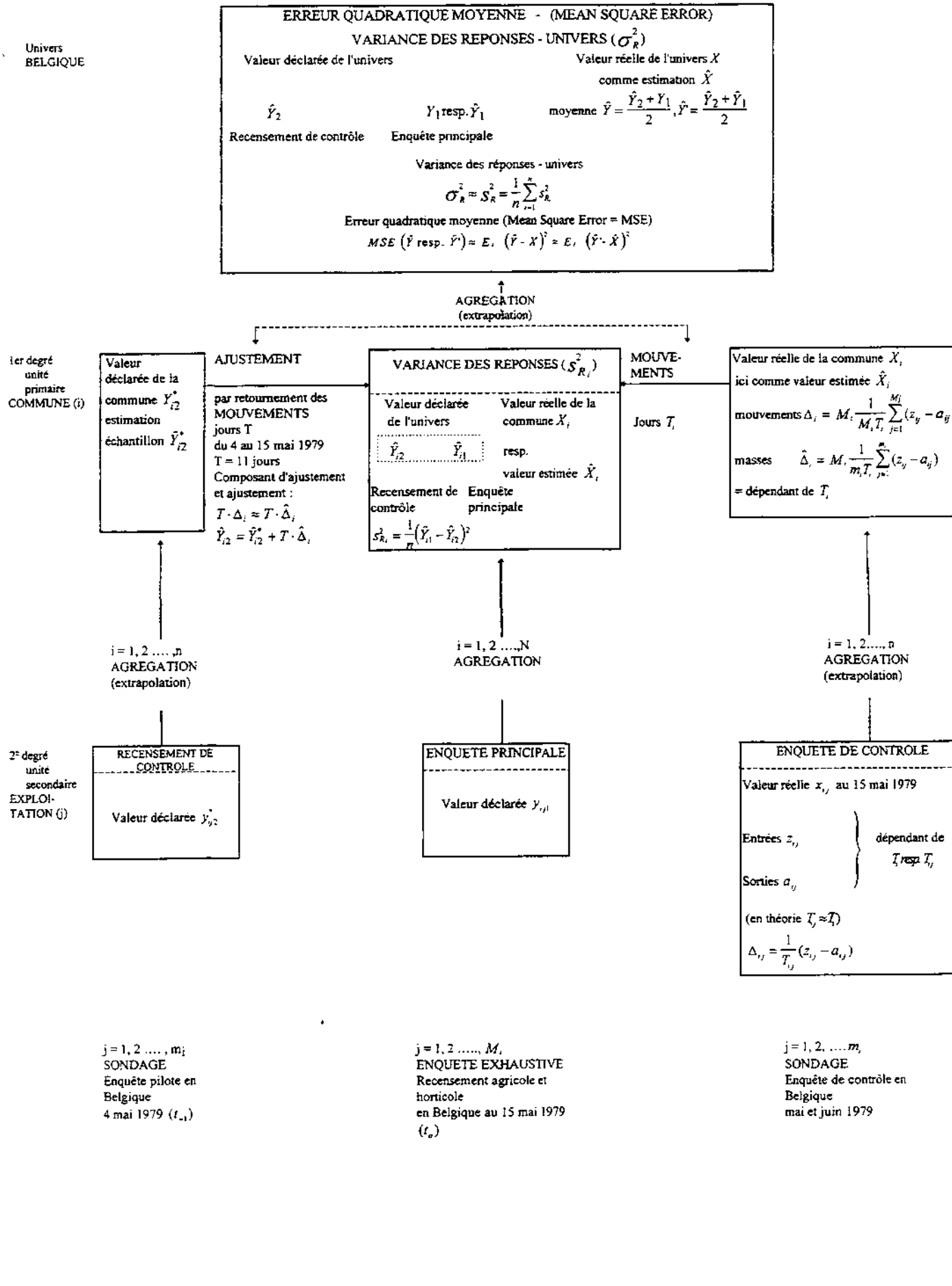
(1) EQM ($\hat{Y}$ resp. $\hat{Y}'$) $\approx E_i(\hat{Y} - X)^2 \approx E_i(\hat{Y}' - X)^2$ ou

(2) EQM $(\bar{y}_{(2)n}) \approx E_i(\bar{y}_{(2)n} - \bar{X})^2 \approx E_i(\bar{y}_{(2)n} - \hat{\bar{X}})^2$.

Dans le schéma, l'on n'a représenté que l'hypothèse (1), modifiée en deux échelons conformément au modèle de sondage. Sans difficulté aucune et moyennant quelques modifications l'on peut aussi représenter l'hypothèse (2). Pour simplifier l'on n'a calculé dans l'hypothèse (2) que l'exemple mathématique du EQM à deux échelons. La valeur relative $\sqrt{\text{EQM}}$ (%) = 4,6 % est la même pour (1) et (2).

SCHEMA
ERREUR QUADRATIQUE MOYENNE ET VARIANCE DES REPONSES

Schéma pour déterminer l'erreur quadratique moyenne et la variance des réponses sur la base de deux recensements (de contrôle) à deux degrés et d'une enquête de contrôle à deux degrés



Exemple de calcul

L'exploitation $j = 1$ dans la commune B-46020 Saint-Gilles-Waës (i) en Flandre-0rientale:
Hausse (z_{ij}) - Baisse (a_{ij}) (Modifications entre le 15 mai 1979 et le jour de contrôle - constatées au moyen de l'enquête de contrôle - ici $T_i = 30$ jours) = 25 - 10 = 15 porcs - modification journalière moyenne du cheptel réel :

$$\Delta_{ij} = \frac{15}{30} = 0{,}50 \text{ porc dans l'exploitation } j$$

valeurs estimées des modifications moyennes du cheptel porcin réel de la commune de Saint-Gilles-Waës (i) :

$$\Delta_i \approx \hat{\Delta}_i = \frac{M_i}{m_i} \sum_{j=1}^{m_i} \Delta_{ij} = \frac{235}{8} \cdot 1{,}40 = 41{,}125 \text{ porcs}$$

Ajustement

$\Delta_1 t = \Delta t = t_O - t_{-1} = 11$ jours, la composante d'ajustement estimée dans cette commune est donc

$T \cdot \hat{\Delta}_i = 11 \cdot 41{,}125 = 452{,}375$ porcs (t_O = 15 mai, t_{-1} = 4 mai)

Valeur déclarée estimée de l'enquête répétée au 4 mai 1979 :

$$\hat{Y}^*_{i2} = \frac{M_i}{m_i} \sum_{j=1}^{m_i} y^*_{ij2} = 118\,939{,}375 \text{ porcs}$$

Valeur déclarée estimée de l'enquête répétée, ajustée au 15 mai 1979 :

$$\hat{Y}_{i2} = \hat{Y}^*_{i2} + T \cdot \hat{\Delta}_i = 119391{,}8 \text{ porcs}$$

Valeur déclarée estimée de l'enquête principale au 15 mai 1979 :

$$Y_{i1} \approx \hat{Y}_{i1} = \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij1} = 118675{,}0 \text{ porcs}$$

Variance des réponses estimée de la commune (i) de Saint-Gilles-Waës

$$\sigma^2_{Ri} \approx s^2_{Ri} = \frac{1}{2}(\hat{Y}_{i1} - \hat{Y}_{i2})^2 = 256865{,}281$$

La variance des réponses de la caractéristique au niveau de la Belgique revient à l'estimation suivante :

$$\sigma_R^2 \approx S_R^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{S_{Ri}^2}{N \cdot P_i} = \frac{1}{129}313083487{,}5 = 2427003{,}779$$

Variance des réponses des valeurs déclarées des communes belges (valeurs déclarées agrégées) dans laquelle :

N = nombre de communes belges (596)

M = nombre d'exploitations en Belgique (44253 au 15 mai 1979 - l'on note à peine quelques changements aux alentours de cette date)

P_i = probabilité de sélection de la commune i comme unité primaire, proportionnellement au nombre de porcs dans la commune

$$\sum_{i=1}^{N} P_i = 1$$

Lorsque le nombre d'exploitations par commune =

$$\overline{M} = \frac{M}{N} = \frac{44253}{596} \approx 74{,}25$$

la valeur estimée des valeurs déclarées des exploitations est :

$$\frac{S_R^2}{\overline{M}^2} = 440{,}22$$

Tableau 2

Recensement agricole et horticole au 15 mai 1979 en Belgique

Définition de la variance des réponses et de la variabilité des réponses avec ajustement; caractéristique : code 365 nombre de porcs

Calculs

Table de calcul a)

Données pour estimer la variance des réponses dans la commune (*i*) B - 46020 Saint-Gilles-Waës (4. Province de Flandre-Orientale)

Probabilité de sélection $P_i = \frac{{}^*Y_i}{{}^*Y} = \frac{33827}{4992276} = 0{,}0067759$

N° de l'exploitation (*j*)	Enquête principale du 15 mai 1979	Enquête répétée - enquête-pilote - du 4 mai 1979	Valeur réelle au 15 mai 1979	Nombre de jours entre le 15 mai 1979 et le contrôle	Modifications entre le 15 mai 1979 et le contrôle	Modifications quotidiennes des cheptels réels
(*j*)	y_{ij1}	y^*_{ij2}	x_{ij}	$T_i(=T_{ij})$	$z_{ij} - a_{ij}$	Δ_{ij}
1	66	55	66	30	15	0,50
2	120	98	132	30	18	0,60
3	298	292	310	30	0	0,00
4	3 326	3 326	3 326	30	- 6	- 0,20
5	62	107	85	30	18	0,60
6	38	50	38	30	- 2	- 0,07
7	99	100	100	30	- 1	- 0,03
8	31	21	32	30	0	0,00
	$\sum_{j=1}^{8} y_{ij1} = 4040$	$\sum_{j=1}^{8} y^*_{ij2} = 4049$				$\sum_{j=1}^{8} \Delta_{ij} = 1{,}40$

(${}^*Y_i$; ${}^*Y$ = nombre de porcs d'après le recensement agricole précédent du 1er décembre 1978).

$M_i = 235$ exploitations dans la commune (*i*) ; $m_i = 8$ exploitations dans la commune (*i*) (dans le recensement répété et l'enquête de contrôle).

z_{ij} = hausse ; a_{ij} = baisse.

Tableau 2 (suite)

Table de calcul b)

Estimation des variances de réponses ($s_{R_i}^2$) dans les communes reprises dans l'échantillon (extrait)

N° de la commune de l'échantillon (i)	Province, commune et code d'après la liste alphabétique des communes	Probabilité de sélection P_i	Nombre d'exploitations M_i	Nombre d'exploitations reprises dans l'échantillon m_i	Enquête principale (valeur estimée) $\hat{Y}_{i1}$	Enquête répétée (valeur estimée) $\hat{Y}_{i2}^{*}$	Composant d'asjustement (valeur estimée) $11.\hat{\Delta}_i$	Valeur estimée ajustée de l'enquête répétée $\hat{Y}_{i2}$	Variance de réponse (valeur estimée) $s_{R_i}^2$	Variance transformée $s_{R_i}^2/(N.P_i)$
	1. Province d'Anvers									
1	11009 Brecht	0,0073742	141	8	35814,0	36395,6	3173,80	39569,4	7051629,329	1604459,633
.	.									
.	.									
7	12040 Willebroek	0,0004781	24	8	3441,0	3216,0	96,10	3312,1	8306,927	29150,116
.	.									
.	.									
.	.									
.	4. Province de Flandre orientale									
.	.									
.	.									
67	46020 Saint-Gilles-Waës	0,0067759	235	8	118675,0	118939,4	452,37	119391,8	256865,281	63605,439
.	.									
.	.									
.	9. Province de Namur									
129	92139 Gembleaux-sur-Orneau	0,001416	68	8	4547,5	5057,5	183,66	5241,2	240582,593	285074,616

$$\sum_{i=1}^{n=129} \frac{s_{Ri}^2}{N.P_i} = 313083487{,}5$$

Remarque : les communes qui, à deux, trois ou plusieurs reprises, se sont retrouvées dans l'échantillon ont été, lors des calculs, prises en considération à deux, trois ou plusieurs reprises. Sur les 129 communes de l'échantillon, 41 communes ont été sélectionnées à plusieurs reprises sur base d'un prélèvement "avec remise"; c'est pourquoi l'échantillon de l'enquête répétée (enquête-pilote) comportait 88 communes différentes.

Pour ce qui est du cheptel porcin réel, les valeurs relatives correspondantes, respectivement $\hat{\bar{X}}_{Nn}$ et $\hat{\bar{X}}_{Mn}$ (provenant d'une enquête de contrôle de l'échantillon), respectivement pour chaque commune et pour chaque exploitation, comportent

$$\frac{S_R}{\hat{\bar{X}}_{Nn}} = \frac{S_R}{\bar{M}\hat{\bar{X}}_{Mn}} = 17{,}8\ \%\ \text{où}\ \hat{\bar{X}}_{Nn} = 8768{,}1\ \text{porcs}$$

$$\text{et}\ \hat{\bar{X}}_{Mn} = \frac{N}{M}.\hat{\bar{X}}_{Nn} = 118{,}1\ \text{porcs}$$

L'on obtient la variabilité des réponses relative ou variance d'échantillonnage des réponses (racine carrée) comme suit : $\frac{S_R}{\hat{\bar{X}}_{Nn}\sqrt{2n}} = \frac{S_R}{\bar{M}\hat{\bar{X}}_{Mn}\sqrt{2n}} = 1{,}11\%$

La valeur 1,11 % est sans doute un peu trop élevée au motif que l'information provenant du recensement exhaustif n'a été reprise dans l'estimation que pour autant que l'on disposait d'informations provenant de l'échantillonnage de l'enquête répétée volontaire (enquête-pilote). Si l'on imagine une information complète en remplaçant *n* par *N* dans la formule de la variance, il en résulte une valeur inférieure, d'environ 0,2 % pour l'expression relative. L'on peut supposer qu'une valeur améliorée, bien qu'en pratique non définissable, se situe dans un intervalle fixé entre 0,2 % et 1,11 % (BECKMANN, WIEGERT, 1987).

Comme l'enquête de contrôle a permis d'établir les valeurs réelles des cheptels porcins dans les exploitations, l'on a pu calculer en tant qu'estimation basée sur l'échantillon l'erreur quadratique moyenne - EQM (mean square error - MSE) comme mesure de l'erreur totale (STRECKER, WIEGERT, PEETERS, etc., 1983) :

$$\hat{\text{EQM}} = 163666{,}86\ \text{ou}\ \sqrt{\hat{\text{EQM}}}(\%) = 4{,}6\%.$$

Si l'on part de l'estimation ponctuelle de la grandeur $\sqrt{\hat{\text{EQM}}}$, l'on constate qu'un quart de la valeur EQM renvoie à l'influence de la variabilité des réponses; il est aussi possible avec le modèle traité ici de donner une estimation de l'EQM (STRECKER, WIEGERT, 1989); il en ressort que la part de la variabilité des réponses dans l'EQM peut atteindre un tiers puisque, en particulier, on peut dire :

En ce qui concerne la variable "nombre de porcs" les valeurs estimées de chaque commune (en divisant cette valeur estimée par la valeur moyenne du nombre d'exploitations par commune - 74,25 - l'on obtient les valeurs estimées correspondantes par exploitation; les valeurs relatives - % - restant inchangées) étaient :

$\hat{B}$ = -320,628 resp. -2,0 (%) biais, ici erreur de déclaration

s_R = 1557,884 resp. 17,8 (%) variance des réponses (racine carrée)

$\frac{s_R}{\sqrt{2n}}$ = $96,989_6$ resp. 1,1 (%) variance et variabilité d'échantillonnage des réponses (racine carrée)

$\sqrt{\text{EQM}}$ = 404,558 resp. 4,6 (%) erreur quadratique moyenne (racine carrée)

Ces chiffres montrent qu'environ un quart de la valeur de l'erreur quadratique moyenne estimée (racine) (estimation ponctuelle) renvoie à l'influence de la variabilité des réponses.
Des résultats correspondants pour l'estimation des estimateurs de l'erreur quadratique moyenne sont :

$$\sqrt{\hat{\text{EQM}}_{Min}} \leq \sqrt{\hat{\text{EQM}}} \leq \sqrt{\hat{\text{EQM}}_{Max}}$$

$$392,76 \leq \sqrt{\hat{\text{EQM}}} \leq 416,02$$

$$4,48\% \leq \sqrt{\hat{\text{EQM}}}\,(\%) \leq 4,74\%$$

La limite inférieure de l'estimation est une valeur de l'erreur quadratique moyenne qui ne tient pas compte de la variance des réponses ($s_R^2 = 0$) . L'estimation de droite représente la limite supérieure de l'influence de la variance des réponses sur l'EQM. Il en résulte l'estimation approximative suivante en ce qui concerne la part de la variance (variabilité) des réponses dans l'estimateur EQM :

$$Q_{Min} := \frac{0}{\sqrt{\hat{\text{EQM}}_{Max}}} \leq Q := \frac{s_R / \sqrt{2n}}{\sqrt{\hat{\text{EQM}}}} \leq Q_{Max} := \frac{s_R / \sqrt{n}}{\sqrt{\hat{\text{EQM}}_{Min}}}$$

$$Q_{Min} = \frac{0}{416,02} = 0 \leq Q \leq Q_{Max} = \frac{137,16}{392,76} \approx \frac{1}{3}.$$

Ou en termes plus simples :

La part de la variabilité des réponses dans l'estimateur de l'erreur quadratique moyenne (racine) peut dans cet exemple atteindre un tiers de la valeur de l'erreur quadratique moyenne même (racine).

Lors de l'estimation de la variabilité des réponses en tenant compte d'un ajustement, il est recommandé d'effectuer aussi tous les calculs sans ajustement pour voir si celui-ci influe fortement les résultats estimés. Lorsque l'état des données le permet, l'on peut aussi effectuer des ajustements en utilisant diverses méthodes afin de pouvoir évaluer l'impact de la méthode d'ajustement. Il est apparu dans l'exemple donné qu'il n'y avait pas de différence fondamentale dans l'ordre de grandeur des résultats (la différence n'était que de 0,14 %). Dans d'autres cas il n'en sera pas ainsi et il sera nécessaire de procéder à de profondes réflexions (STRECKER, WIEGERT, PEETERS, etc., 1983).

En résumé, il ressort de la première partie de notre texte que définir la variabilité des réponses représente une contribution importante dans l'amélioration de la qualité des données et donc de la fiabilité de la statistique. Les modèles d'enquêtes répétées qui ont été développés à cette fin portent l'empreinte de l'ajustement lorsque pour des caractéristiques non-constantes dans le temps il est souhaitable de convertir différents jours de sondage en une journée uniforme de sondage. Cette façon de travailler éclaire un nouveau facteur source d'erreur - auquel jusqu'à présent l'on avait peu prêté attention - provenant de la longueur de la durée de l'enquête. Elle est essentielle lorsque les valeurs des caractéristiques changent avec le temps; elle est, par contre, insignifiante lorsqu'elles sont constantes dans le temps. En tout cas, l'on ne peut négliger, d'un point de vue organisationnel, de prévoir une durée d'enquête aussi réduite que possible afin de restreindre au maximum l'influence de ce facteur dans l'ensemble de toutes les autres erreurs de données.

V. Conclusion de la première partie

Pour conclure, résumons à nouveau les points principaux.

L'enquête répétée et ses divers développements (première partie II C) forment une méthode permettant d'estimer des variabilités de réponses individuelles survenant lors d'enquêtes dans lesquelles les personnes interrogées remplissent elles-mêmes les questionnaires. Au sein du concept enquête répétée, l'ajustement dans le temps des données provenant d'enquêtes effectuées *à différents jours* est essentiel. C'est la raison pour laquelle l'on a exposé dans cette partie de quelles diverses façons l'on peut effectuer un ajustement - de manière exacte ou le plus souvent seulement comme estimation - et quels en sont les résultats. Deux exemples pratiques accompagnés des calculs nécessaires en ont fait la démonstration. D'après les statisticiens spécialisés en la matière, l'ordre de grandeur des estimations de variabilités obtenues de la sorte et, par suite, de la qualité des données de l'enquête est plausible et réaliste.

SECONDE PARTIE

VARIABILITE DES REPONSES ET SONDAGES STRATIFIES

I. Modèle d'enquêtes répétées avec stratification

A. Introduction

Beaucoup de résultats d'enquêtes statistiques sont présentés par classes de grandeur (par exemple, nombre d'habitants par classes de grandeur des communes, nombre d'exploitations agricoles par classes de grandeur de la superficie agricole, nombre d'entreprises industrielles par classes de grandeur de l'activité). En pratique aussi l'on trouve beaucoup de sondages stratifiés (ces strates sont d'ailleurs souvent des classes de grandeur). Vu l'importance des modèles d'erreurs d'enquête et de réponse, ceux-ci et les problèmes liés seront approfondis ci-après.

Dans le cas présent sera étudié un sondage stratifié par classes de grandeur avec enquête principale et enquête répétée à caractéristique d'enquête constante dans le temps (à tout le moins pour la durée des enquêtes). Les classes de grandeur (strates) peuvent être définies au préalable par l'objectif économique des résultats statistiques : classes de grandeur "administratives" ou économiques ou stratification "administrative" ou économique sans limitation optimale des strates, c'est-à-dire sans stratification optimale (pas de strates "techniques") selon le sens défini par Dalenius (DALENIUS, 1951, 1957; COCHRAN, 1977; SCHNEEBERGER, 1985, 1991).

Il faut adapter l'ensemble des formules exposées précédemment pour les enquêtes répétées sous la forme de sondages sans stratification (cf. la première partie de cet article) aux conditions modifiées de la stratification afin de pouvoir aussi, par analogie avec ce qui précède, calculer dans ce cas des estimations de la variance des réponses. Dans la seconde partie donc ces formules seront développées. En outre, l'on trouvera dans la section II de cette même seconde partie l'examen d'une problématique particulière qui se produit lors de l'application de sondages stratifiés à l'estimation des tendances principales et des résultats en résultant.

Dans le cas d'enquêtes stratifiées - principales et répétées - il peut en effet se produire que suite à la variabilité des réponses les valeurs enquêtées d'une même unité concernent différentes classes de grandeur. Comment faut-il tenir compte de ce phénomène et dans un tel contexte quelle peut être la corrélation entre les effectifs dans les diverses classes de grandeur en particulier et la fiabilité de l'ensemble de la répartition en classes ? Cette question montre que la variabilité des réponses peut provoquer une inconsistance spécifique pour les résultats répartis en classes; nous avons analysé ce point dans la section II de la présente 2e partie.

Dans la section III nous présentons des estimations de la variabilité des réponses dans les déclarations de superficie, basées sur les résultats du Recensement agricole et horticole du 15 mai 1985 en Belgique (dans lequel on a pris comme enquête principale un sondage stratifié par classes de grandeur) et sur les résultats d'une enquête répétée (enquête-pilote; sondage du 4 mai 1985 stratifié d'après les mêmes classes de grandeur, enquête répétée ex ante) pour la catégorie 1. D'expérience nous savons qu'une caractéristique - comme la surface cultivée de toute exploitation durant la période de végétation et en particulier pendant le mois de mai - reste constante durant la période de végétation à tout le moins pendant une courte durée. Les achats et ventes effectués par l'ensemble des exploitations durant les quinze jours entre l'enquête principale et l'enquête répétée en mai 1985 étaient tellement faibles (seulement 0,6 ‰ du total des terres agricoles ont été échangées en moyenne entre les deux enquêtes) que la caractéristique surface cultivée par exploitation peut être considérée comme constante. Les changements intervenus n'exerçaient donc pas d'influence mesurable sur la variabilité des réponses. S'il n'est donc pas nécessaire de procéder à un ajustement, il faut par contre tenir compte de la stratification et des limitations que la variabilité des réponses impose à celle-ci. C'est pourquoi dans la section III de cette 2e partie l'on a ajouté aux estimations proposées des recherches sur les déclarations de consistance et d'inconsistance. L'exposé se termine par une évaluation de l'ensemble des résultats.

En anticipant sur le développement futur des enquêtes relatives à la superficie agricole, l'on peut avancer qu'une partie des déclarations sera collectée au moyen de méthodes statistiques de télédétection (vues aériennes) et non plus en questionnant les exploitations. Du coup, les variabilités dues au questionnement tombent et l'on obtiendra probablement de la sorte une valeur approchée améliorée de la superficie utilisée à condition que des erreurs d'appréciation de même taille ne remplacent les erreurs de réponse.

Nous savons déjà que cette méthode d'enquête objective de télédétection (mesurage au lieu de questionnement), même si le système de travail est très correctement effectué, provoque des écarts et des distorsions dans les résultats (voir RADERMARCHER 1990, 1992 a, 1992 b; ainsi que STADLER 1989, 1991). Pour déterminer ce genre d'erreurs, il faut développer d'autres modèles que ceux que nous proposons ici.

B. Estimation de la variabilité des réponses en cas de stratification

Dans une enquête à L strates, chaque strate peut être conçue comme l'application d'un modèle à choix aléatoire indéfini. Dans la section II B de la première partie, nous avons parlé en détail du modèle d'erreur de réponse en cas de sondage aléatoire indéfini. Les formules que nous y avons avancées peuvent, en cas de stratification, être appliquées sans modification pour effectuer les calculs dans chaque strate $h = 1,2,..,L$. Par suite, elles ont été pourvues d'un indice h. L'agrégation importante au niveau de toutes les strates donne les formules pour l'estimation totale. Il ressort :

$$s_R^2 = \frac{1}{N}\sum_{h=1}^{L} N_h \cdot s_{R_h}^2 \quad \text{avec } N = \sum_{h=1}^{L} N_h$$

Variance des réponses dans la strate h (estimation)

Valeur estimée de la variance des réponses totales

$$\hat{V}_R = \frac{1}{N^2}\sum_{h=1}^{L} N_h^2 \cdot \hat{V}_{R_h}$$

Valeur estimée de la variabilité des réponses totales

où

$$\hat{V}_{Rh} = \frac{s_{Rh}^2}{n_h \cdot k}$$

Variabilité des réponses dans la strate h (estimation)

Lorsqu'une strate a été exhaustivement ou entièrement soumise à enquête, il faut mettre à la place de $s_{Rh}^2 = \frac{1}{n_h}\sum_{i=1}^{n_h} s_{R_i h}^2$ la variance de réponses $S_{Rh}^2 = \frac{1}{N_h}\sum_{i=1}^{N_h} s_{R_i h}^2$.

II. Effet de la variabilité des réponses sur les résultats d'enquêtes stratifiées

Comme nous l'avons déjà souligné à plusieurs reprises, une enquête statistique sera presque toujours, en pratique, une enquête unique, non répétitive. L'on n'obtiendra, ce faisant, qu'une seule valeur observée de la caractéristique de chaque unité *i*. Les enquêtes répétées dont question dans la première partie de ce texte prévoient des examens multiples (objet et moment identiques) pour rendre mesurable l'influence de certaines erreurs sur les données. Elles forment donc un instrument de contrôle de l'exactitude des données examinées par rapport aux erreurs de déclaration. Pour les raisons déjà évoquées, l'on se limite à tout au plus deux de ces "enquêtes répétées". Si celles-ci sont stratifiées, leurs résultats pourront être utilisés non seulement pour estimer la variabilité des réponses des données examinées, mais aussi, en outre, pour contrôler la fiabilité de la répartition des fréquences en classes, obtenue à partir de l'enquête. En procédant de la sorte, il n'est pas sûr que les deux résultats d'enquête tombent dans des strates identiques. Ci-après, cette problématique sera expliquée au moyen d'une enquête principale et d'une enquête répétée avec stratification administrative ou économique d'après un même système de travail. La variable de l'enquête est en même temps la variable de la stratification.

Pour l'unité d'enquête d'indice *i*, l'on obtient donc deux valeurs examinées y_{i1}, y_{i2}, qui forment une information équivalente pour attribuer l'unité et sa caractéristique à une strate.
Deux cas distincts peuvent se produire :
1. y_{i1}, y_{i2} tombent dans la même strate ou classe de grandeur;
2. y_{i1}, y_{i2} appartiennent à différentes strates (classes de grandeur).

Le premier cas ne pose pas problème. Dans le deuxième, par contre, il faut décider à quelle strate (classe de grandeur) attribuer l'unité *i*. Si l'on considère y_{i1}, y_{i2} comme équivalentes eu égard à leur précision statistique, on définit $(y_{i1} + y_{i2})/2$ comme la moyenne de la valeur pour laquelle l'erreur aléatoire est réduite et on range l'unité *i* correspondante dans la classe de grandeur appropriée afin d'établir de la sorte une "répartition des fréquences moyenne".

La compensation, qui s'établit ainsi entre les répartitions des fréquences de chaque enquête, permet d'exploiter les résultats incertains d'une seule enquête au moyen d'informations classées équivalentes, doublement disponibles, et d'avancer une valeur moyenne de l'inconsistance (*Index of Inconsistency*) de la répartition des fréquences en résultant.

Mais il ne faut pas perdre de vue que l'on n'obtient que rarement des résultats provenant d'enquêtes répétées permettant de telles vérifications. En vue d'assurer la comparabilité avec la pratique statistique qui ne prévoit à chaque fois qu'une seule enquête, il semble justifié d'un point de vue pragmatique, même si l'on dispose d'une double information, de ne pas classer d'après la valeur moyenne issue des deux enquêtes mais d'après la valeur de l'enquête principale. Ce pragmatisme n'enlève rien aux considérations sur l'inconsistance de la répartition des fréquences.

L'on peut mesurer l'incertitude liée à l'attribution des déclarations individuelles, provenant d'un seul recensement - l'enquête principale - comme c'est la règle dans la pratique statistique, au moyen d'informations d'attribution équivalentes ($k \geq 2$) qui ont, par exemple, été collectées à partir de l'enquête d'origine ou principale ($t = 1$) et ($k = 1$) ≥ 2 enquêtes répétées ou au moins un recensement répété ($t = 2$). Comme mesure moyenne totale de l'inconsistance existante des valeurs individuelles, l'on peut utiliser l'indice d'inconsistance individuel (*Index of Unreliability* ou *Index of Inconsistency*), développé par le U.S. Bureau of the Census (HANSEN, HURWITZ, PRITZKER, 1964), à condition toutefois de l'élargir pour pouvoir l'appliquer à L classes de grandeur (STRECKER, 1997 et 1999).

Avant de calculer l'indice, il faut, à l'aide des variables (0 ; 1), rendre dichotomiques les réponses individuelles $y_{i1G}, y_{i2G}, \ldots y_{ikG}$, des unités i pour i = 1, 2....., N resp. n.

Les valeurs déclarées individuelles y_{it} resp. y_{itG} ont été collectées selon une méthode de travail G identique ou une méthode G' (G ≈ G') analogue.

Soit Z_{itG} : Z_{i1G}, Z_{i2G}, ..., Z_{ikG} (0 ; 1) variables de l'unité i.

Celles-ci sont définies comme suit :

$$Z_{itG} = \begin{cases} 1, \text{ lorsque la réponse donnée } y_{itG} \text{ de l'unité i indique pour une caractéristique une propriété A dans le recensement répété t, il faut ici attribuer l'unité i à la classe de grandeur h.} \\ 0, \text{ lorsque la réponse donnée } y_{itG} \text{ de l'unité i montre une autre caractéristique (pas A, en symbole } \overline{A}\text{) l'unité i en ce cas ne doit pas être attribuée à la classe de grandeur h.} \end{cases}$$

L'indice d'inconsistance est maintenant défini sous la forme $I_{RGh} = \frac{\sigma_R^2}{P_G(1-P_G)}$;

où $\sigma_R^2 = E_i\left[E_t\left(Z_{itG} - E_t(Z_{itG})\right)^2\right]$ est la variance de réponses individuelle moyenne = variance de réponses simple et $\sigma_{R_i}^2$ = variance de réponses individuelle de l'unité i.

$P_G = E_i(P_{iG}) = E_i\left(E_t(Z_{itG})\right) = E_{i,t}(Z_{itG})$

= la moyenne -univers total des variables rendues dichotomiques dans l'univers,

$P_G \cdot (1 - P_G)$ = variance d'une variable aléatoire binomiale (0 ; 1).

L'indice est un critère indiqué pour estimer la fiabilité des déclarations des unités i au motif qu'elles peuvent se prévaloir de la caractéristique "appartenant à la classe de grandeur h".

L'intervalle de valeur est :

$$0 \leq I_{RGh} \leq 1.$$

Si les résultats des recensements répétés n'indiquent pas de variabilité des réponses par rapport à l'insertion d'unités i dans les classes de grandeur h, l'indice a la valeur 0; en cas d'incertitude totale (différences "aléatoires") des réponses l'on obtient une valeur de l'indice 1.

Comme pour des raisons organisationnelles que nous avons déjà évoquées, l'on n'exécute en pratique qu'une seule enquête principale et une seule enquête répétée (le recensement répété comme échantillon ou sous-échantillon), la formule d'estimation de l'indice pour un échantillon de n unités i est donnée dans ce cas pour un échantillon illimité avec répartition bidimensionnelle a posteriori des unités par classes de grandeur (table de contingence) (pour plus d'informations, voir STRECKER, 1997).

Pour effectuer ce calcul dans le cas d'un échantillon avec n unités l'on part du schéma suivant :

Schéma à quatre cases pour la classe de grandeur h

	Enquête principale ou originale (t=1) Caractéristique :		
Recensement répété (t resp. t'=2)	A $z_{i1G} = 1$	$\overline{A}$ $z_{i1G} = 0$	$\sum$
Caractéristique A $z_{i2G'} = 1$	a	b	$a + b$
$\overline{A}$ $z_{i2G'} = 0$	c	d	$c + d$
$\sum$	$a + c$	$b + d$	n

Les valeurs a, b, c et d dans ce tableau sont des paires absolues que l'on trouve fréquemment dans les deux enquêtes (nombres naturels y compris 0) (t=1, t resp. t'=2, G=G' resp. G≈G') et n=a+b+c+d. Après quelques transformations l'on obtient comme valeur estimée pour

$$\sigma_R^2 \approx \frac{b+c}{2n} \quad , \; P_G(1-P_G) \approx \frac{1}{2}\left[\left(\frac{a+c}{n}\right)\left(\frac{b+d}{n}\right)+\left(\frac{a+b}{n}\right)\left(\frac{c+d}{n}\right)\right]$$

et donc pour la formule d'estimation de l'indice d'inconsistance

$$\hat{I}_{RGh} = \frac{(b+c)}{\frac{(a+c)(b+d)+(a+b)(c+d)}{n}}$$

L'intervalle des valeurs de l'estimateur $0 \leq \hat{I}_{RGh} \leq 1$ est uniquement valable dans les conditions $b^2 + c^2 \leq 2ad$, qui, en pratique, sont généralement satisfaites (cf. GROVES, 1989, et KANTOROWITZ, 1969).

Concernant le classement erroné (*misclassification*) dont on a parlé ici, l'indice indique en quelle mesure les résultats de la première et de la seconde enquête se recoupent ou varient d'après leur classement. Lors d'enquêtes stratifiées comme dans le cas présent l'on obtient pour chaque strate ou classe de grandeur h une valeur d'indice $\hat{I}_{RGh}$, avec h = 1, 2..., L; une moyenne géométriquement pondérée ou mathématique les réunit dans un indice total (voir à ce sujet le tableau 6b).

Le test de fiabilité préalable ne pouvait valoir que provisoirement car l'élaboration correcte d'une répartition de probabilité répartie en classes suppose de connaître les réelles valeurs y_i individuelles.

III. Enquête répétée dans le cas de caractéristiques constantes dans le temps. Résultats d'une enquête en Belgique

Nous allons maintenant, pour illustrer les estimations dont nous parlions dans la première partie, donner les résultats de l'estimation de la variabilité des réponses lors d'une enquête stratifiée en Belgique, le Recensement agricole et horticole au 15 mai 1985 - catégorie 1. Cette catégorie comprend les agriculteurs et éleveurs dont c'est la profession principale. On entend, par là, l'occupation qui absorbe la plus grande partie du temps ou, dans les cas douteux, celle qui procure le revenu le plus important (voir le questionnaire dudit recensement - modèle 1).

Dans ce cas précis, comme l'on avait fixé à l'avance la manière d'utiliser des échantillons stratifiés à caractère économico-administratif, il a fallu en tenir compte en établissant le planning, ainsi que lors du calcul des résultats et lors de l'estimation. Dans la première partie de cette étude, l'on a déjà traité de l'estimation de la variabilité des réponses, pour ce qui est de la catégorie 5, sous la forme d'un modèle théorique d'erreur de réponse (enquête principale et enquête répétée exhaustive).

Dans l'introduction nous avions déjà observé que des caractéristiques telles que les superficies d'une exploitation - par exemple, la superficie agricole utilisée, les superficies cultivées de froment d'hiver, orge d'hiver, betteraves et plantes industrielles - peuvent être considérées, sur une période assez longue (une période de végétation, par exemple), comme présentant une valeur de caractéristique constante.

A cause du transfert dans le temps entre enquête principale et enquête répétée, cette constance acquiert une signification essentielle dans le cas d'enquêtes répétées. Lors du Recensement agricole belge qui s'effectue périodiquement (chaque semestre), l'on a pu, par exemple au 15 mai 1985, déterminer, en ce qui concerne la catégorie 1 (agriculteurs et éleveurs dont c'est la profession principale), au moyen d'une enquête (questionnaire à compléter par l'exploitant) les superficies par exploitation et les traiter par classes de grandeur de la superficie agricole utilisée. Pour un tel mode d'enquête la variance des réponses et la variabilité des réponses doivent être définies par strate spécifique. Comme enquête principale l'on a choisi le recensement du 15 mai 1985 (catégorie 1) et on l'a reliée à la deuxième enquête, l'enquête-pilote du 4 mai 1985, qui est un sondage stratifié.

L'on a obtenu une taille effective de l'échantillon sur la base des unités dont la valeur apparaissait tant dans l'enquête principale que dans l'échantillon. L'enquête-pilote a été effectuée de manière analogue à l'enquête principale, mais stratifiée d'après les classes de grandeur de l'enquête principale (stratification administrative ou économique) et réalisée sur une base volontaire impliquant que le sujet remplisse lui-même le questionnaire.

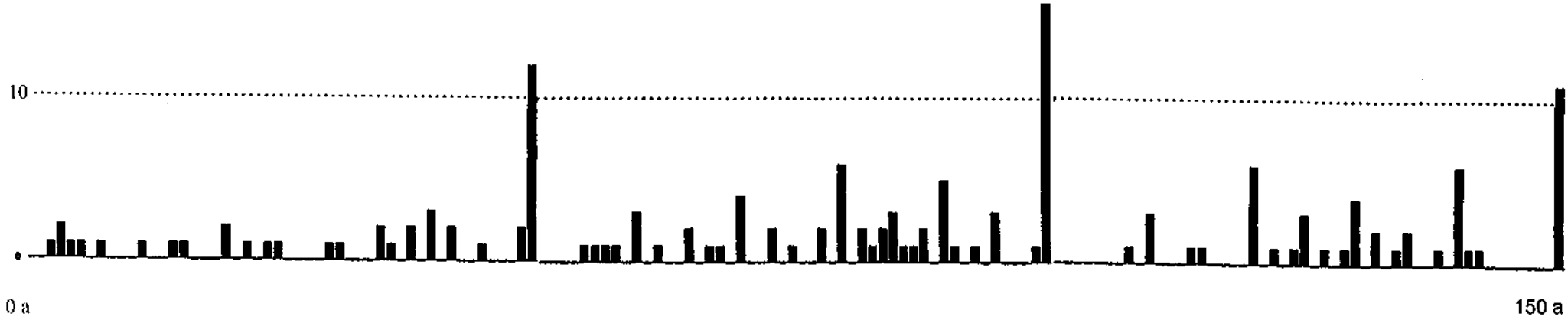
Les orientations décisives des systèmes de travail des deux enquêtes restent de la sorte identiques. Ce point de départ a été conforté a posteriori par les distributions fréquentielles fortement parallèles des données soumises à l'enquête. En ce qui concerne le recensement, l'échantillon tiré de celui-ci d'après le modèle de l'enquête-pilote et d'après l'enquête-pilote elle-même, l'on a forgé de telles distributions fréquentielles des exploitations pour la caractéristique superficie agricole utilisée; cf. les graphiques 1a, 1b et 2 (pages suivantes). Ce qui est remarquable dans ces distributions fréquentielles est justement que l'on observe une corrélation plus qu'évidente dans le champ des erreurs (effet des nombres arrondis au 0 ou 5 supérieur). Pour les valeurs déclarées de la superficie agricole (en ares) se terminant par 0 et 5, c'est-à-dire 5, 10, 15, 20, ..., 50, ..., 100, ... l'on observe des déclarations gonflées. Cela se remarque clairement sur les graphiques 1a, 1b et 2. Dans les graphiques 1 a est l'unité de l'abscisse et 1 exploitation l'unité de l'ordonnée. L'unité are (a) comprend 100 m2, l'unité hectare (ha) 100 ares.
La variable stratifiée, la superficie agricole utilisée par exploitation était répartie en classes de grandeur (strates économico-administratives) comme suit :

Strate h	Superficie agricole (SA) par exploitation en ha
1	0,01 – <10
2	10 – < 30
3	30 – < 100
L = 4	100 – < +

Le Recensement agricole et horticole du 15 mai 1983 a été pris comme base de l'échantillon, l'unité choisie étant l'exploitation de catégorie 1. Des raisons organisationnelles ont fait que le choix devait s'effectuer au plus tard en février 1985.

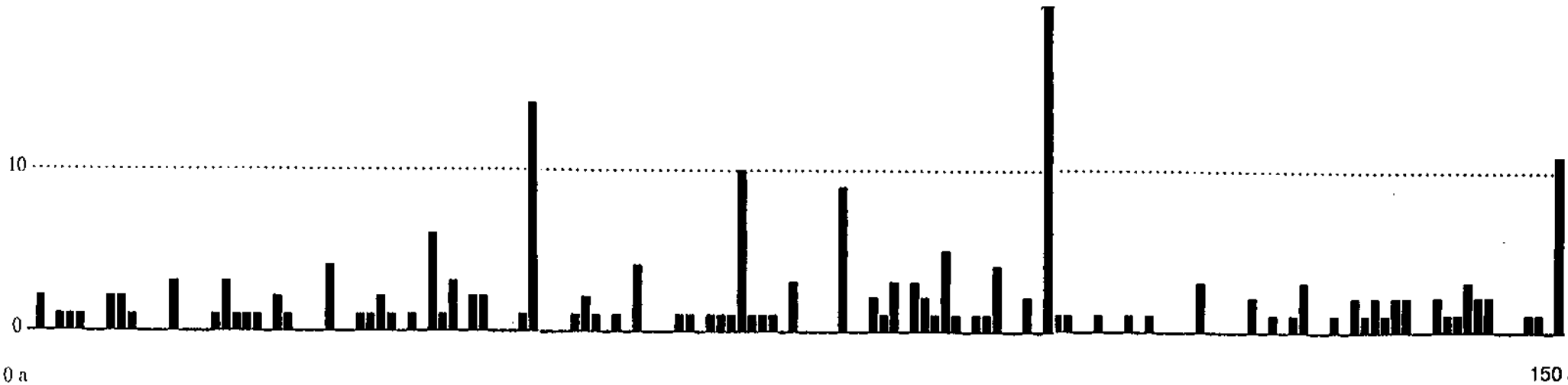
A ce moment-là, on ne disposait pas encore d'une base plus actuelle. La taille de l'échantillon a été calculée avec la précision définie au départ (erreur d'échantillonnage aléatoire - racine simple) de 1 - 2 % des caractéristiques principales : la superficie agricole utilisée, combinée au froment d'hiver. De la sorte la taille de l'échantillon a été portée à n = 2200 exploitations. Celles-ci ont ensuite été réparties entre les strates pour partie de manière proportionnelle et pour partie de manière optimale. Le nombre total d'exploitations de cette catégorie était, d'après le recensement de mai 1985, de 61321, de sorte que l'on obtenait grosso modo un quotient de $\approx$ 3,6 % (2200/61321) (cf. tableau 5d).

Belgique - Recensement agricole et horticole au 15 mai 1985
Catégorie 1 (classe 0 - 150 a)
Enquête-pilote, sondage du 4 mai 1985



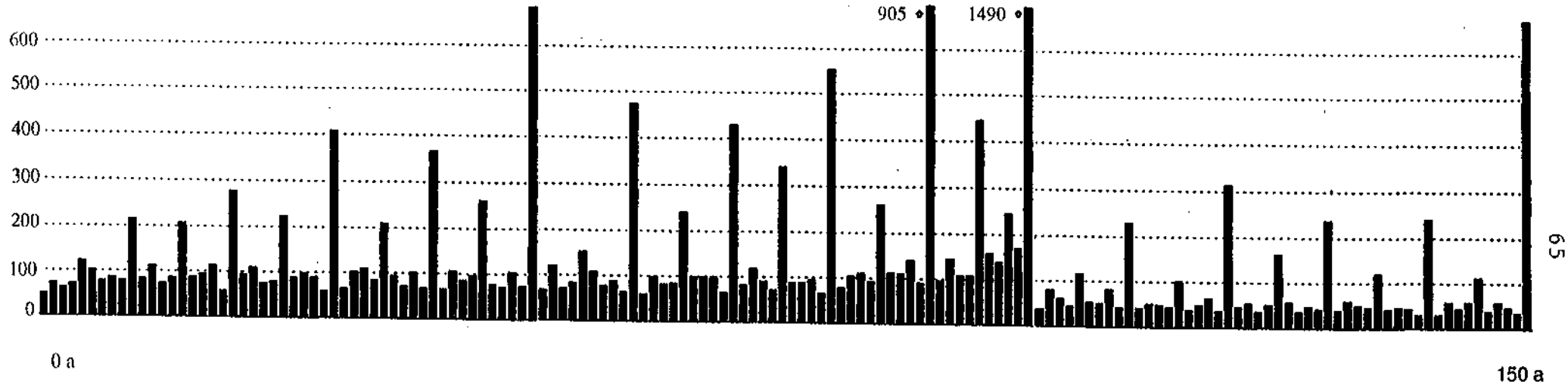
Graphique 1 a : Répartition des exploitations selon la superficie agricole utilisée - enquête-pilote, sondage du 4 mai 1985 (classe 0 - 150 a)

Belgique - Recensement agricole et horticole au 15 mai 1985
Catégorie 1 (classe 0 - 150 a)
Sondage de l'enquête principale du 15 mai 1985



Graphique 1 b : Répartition des exploitations selon la superficie agricole utilisée - sondage de l'enquête principale du 15 mai 1985 (classe 0 - 150 a)

Belgique - Recensement agricole et horticole au 15 mai 1985
Catégorie 1 (classe 0 - 150 a)
Recensement complet du 15 mai 1985



Graphique 2 : Répartition des exploitations selon la superficie agricole utilisée
(classe 0 - 150 a)
Recensement complet du 15 mai 1985

Prendre comme base de l'échantillon l'année 1983 a posé quelques problèmes. Entre 1983 et 198 certaines exploitations avaient cessé leur activité et d'autres étaient venues s'ajouter. Par "nouveau venus" l'on entendait toutes les exploitations qui étaient venues s'ajouter entre le recensement du 15 m 1983 et les 4, respectivement 15 mai 1985; l'"ancien fichier" comprenait les exploitations reprises dans le recensements des 15 mai 1983 et 15 mai 1985. Les nouvelles exploitations avaient généralement é signalées à l'Institut national de statistique par des agents communaux. Tous ces "nouveaux venus" ont é repris dans l'enquête, c'est-à-dire soumis complètement à l'enquête et intégrés en tant que stra complémentaire aux estimations ultérieures. Cela revenait à ce qu'à chaque fois chaque strate devait êt répartie entre une strate "ancien fichier" (AF) et une strate "nouveaux venus" ou "nouveau fichier" (NF L'estimation de la variance des réponses a été de ce fait effectuée séparément dans chaque strate selon A et NF; une moyenne pondérée a permis de réunir les deux estimations. Les particularités de ce mode d calcul sont reproduites dans les tableaux 5 a à 5 e à la fin de la 2e partie.

Les estimations des variances et des variabilités moyennes de réponse que l'on y trouve subissent accessoirement des effets de sondage provenant (contrairement à la catégorie 5) du choix aléatoire de l'échantillon des exploitations dans l'ensemble de la catégorie 1.

Nous avons volontairement renoncé à l'estimation complémentaire de cette composante de l'erreu aléatoire, c'est-à-dire l'estimation des variances de variances, parce que leur impact sur la taille de échantillons disponibles est pratiquement négligeable (cf. BECKMANN, WIEGERT, 1987, pp. 15 160). De plus, les valeurs estimées ont probablement été légèrement surévaluées vu que l'informatio provenant de l'enquête principale (recensement) n'a été reprise dans les estimations que pour autant qu'el fût aussi disponible comme information provenant de l'enquête répétée (enquête-pilote). Dans les tableau 3a, 3b et 4 - à la fin de cette article - l'on trouve les résultats des estimations détaillées de la variance et d la variabilité des réponses des caractéristiques de chaque exploitation :

- superficie agricole utilisée
- superficie froment d'hiver
- superficie orge d'hiver
- superficie betteraves
- superficie plantes industrielles.

Les résultats des variances et des variabilités des réponses dans les tableaux sont des estimations qui, comme nous l'avons déjà dit, proviennent de deux enquêtes, l'une au 4 mai 1985 (enquête-pilote, enquête répétée ex ante) et l'autre au 15 mai 1985 (enquête principale). Vu le caractère constant des caractéristiques de cet intervalle, il n'a pas été nécessaire d'ajuster les résultats de l'enquête-pilote au moment de l'enquête principale du 15 mai 1985. De la sorte, les estimations ne sont pas influencées par des erreurs qui auraient pu être occasionnées par un ajustement. D'autre part, l'on trouve dans les estimations des données provenant de deux enquêtes stratifiées qui, par des classifications erronées (voir à ce sujet la partie II section 2), auraient pu nuire à la qualité de l'estimation. Plus loin, l'on donnera les valeurs de l'indice d'inconsistance qui indique une influence relativement minime de cette source d'erreurs.

Les tableaux 3b et 4 donnent les ordres de grandeur des variabilités de réponses.

C'est dans la caractéristique superficie agricole utilisée par exploitation que celui-ci est avec 0,2 % le plus bas, et le plus élevé avec 0,4 % dans la caractéristique superficie betteraves. La valeur de la variabilité des réponses décroît des classes de grandeur inférieures vers les plus élevées. Cela peut être lié au fait que les exploitations plus importantes sont organisées de façon déjà plus bureaucratique que les petites. En moyenne, la valeur avoisine 0,5 % , un résultat qui indique que les conséquences des erreurs de déclaration - exprimées en variabilité des réponses - sur la qualité des déclarations de superficie dans le recensement agricole et horticole sont relativement limitées. Pour les enquêtes qui ont lieu périodiquement comme le recensement agricole semestriel un tel niveau de qualité est significatif et en fait indispensable; l'investissement important en temps et argent que nécessite la production de telles données statistiques doit aussi se justifier par le contrôle de la qualité des données.

Ces dernières années, dans beaucoup de pays d'Europe occidentale, la technique du questionnement volontaire a été contrariée par des erreurs de non-réponse. Celles-ci se sont aussi fait jour en Belgique lors de l'enquête-pilote qui était une enquête volontaire sans obligation légale. Pour contrôler l'effet de la non-réponse, l'on a déterminé pour chaque strate le taux de non-réponse. L'on a trouvé les valeurs suivantes pour les classes de grandeur 1 à 4 : 11,0 %, 9,3 %, 15,6 % et 8,9 %, soit en moyenne 11,5 %. Ces valeurs sont acceptables et ne remettent pas en cause le fonctionnement de cette méthode d'enquête; elles sont sensiblement inférieures à celles auxquelles a dû faire face l'Office statistique allemand.

Pour ce qui est des considérations en matière d'inconsistance des résultats classés, l'on a établi de tableaux de consistance, adapté à la définition de l'indice, pour chacune des classes de grandeur h = 1, 2 3, 4; ceux-ci sont empruntés au tableau de contingence (universel) pour la répartition des exploitation d'après les valeurs de déclaration de la superficie agricole utilisée dans l'enquête principale (15 mai 1985 et dans l'enquête-pilote (4 mai 1985) selon les classes de grandeur de la superficie agricole utilisé (tableau 6a). En guise d'exemple, l'on trouvera ici le tableau de contingence pour la première classe d grandeur (h = 1), ancien fichier (AF) et nouveau fichier (NF). Des tableaux analogues existent pour le quatre classes au niveau des fichiers anciens et nouveaux.

Tableau de contingence pour la première classe de grandeur (h = 1)

0,01 - < 10 ha de superficie agricole utilisée

a) Ancien fichier (AF)

Enquête principale y_{i1G} / Enquête pilote $y_{i2G'}$	h=1 A $z_{i1G} = 1$	Reste $\overline{A}$ $z_{i1G} = 0$	$\sum$
h=1 A $z_{i2G'} = 1$	$a = 410$	$b = 36$	$a+b = 446$
Reste $\overline{A}$ $z_{i2G'} = 0$	$c = 13$	$d = 1414$	$c+d = 1427$
$\sum$	$a+c = 423$	$b+d = 1450$	$n_{AF} = a+b+c+d =$ 1873

b) Nouveaux venus (NF)

Enquête principale y_{i1G} / Enquête pilote $y_{i2G'}$	h=1 A $z_{i1G} = 1$	Reste $\overline{A}$ $z_{i1G'} = 0$	$\sum$
h=1 A $z_{i2G'} = 1$	$a = 213$	$b = 9$	$a+b = 222$
Reste $\overline{A}$ $z_{i2G'} = 0$	$c = 9$	$d = 246$	$c + d = 255$
$\sum$	$a+c = 222$	$b+d = 255$	$N_{NF} = a+b+c+d =$ 477

h = 1 : 0,01 - < 10 ha ; Reste : 10 ha - < +

$A = 1$ pour $0{,}01 \leq y_{i..} < 10$ ha
$\overline{A} = 0$ pour - 10 ha $\leq y_{i..} <$ +

} superficie agricole utilisée

L'estimation de l'indice pour la première classe de grandeur h = 1 (0,01 - < 10 ha de superficie agricole utilisée), ventilée en ancien fichier et nouveaux venus, en résulte

$$\hat{I}_{RAFh=1} = \left\{ \frac{(b+c)}{\underline{(a+c)(b+d)+(a+b)(c+d)}} \right\} \left\{ = \frac{49}{\underline{423 \,.\, 1450+446 \,.\, 1427}} = 0{,}0734 \approx 7{,}34\% \right.$$

$$I_{RNFh=1} = \left. = \frac{18}{\underline{222 \,.\, 255+222 \,.\, 255}} = 0{,}07583 \approx 7{,}58\% \right.$$

(pour n il faut utiliser tantôt n_{AF} tantôt N_{NF}

L'indice pour la classe de grandeur h = 1 (ancien fichier + nouveaux venus) s'obtient en calculant une moyenne géométrique pondérée (g), les poids étant $N_{AFh=1}$ = 19 017 et $N_{NFh=1}$ = 222 exploitations (voir tableau 4a) , et l'on parvient à une valeur estimée de :

$\hat{I}_{gh=1}$ = 7,3 %

De manière analogue, l'on a établi les tableaux de contingence des autres classes de grandeur et calculé les indices. Les valeurs des indices pour les classes 1 à 4 sont : 7,3 %, 10,0 %, 7,2 % et 8,9 %.

L'indice total est de 8,5 %.

Le tableau 6b détaille les valeurs des indices. Le calcul de l'indice d'inconsistance nécessite théoriquement un échantillon aléatoire illimité dont le résultat sera réparti a posteriori dans les classes de grandeur données (strates) conformément aux variables stratifiées (voir à ce sujet les explications ci-dessus, de même que COCHRAN, 1977, et GROBRAS, 1987, concernant la stratification a posteriori). Comme l'on a travaillé ici sur la base d'une stratification a priori, c'est-à-dire un choix aléatoire réduit, cela a pu desservir l'estimation de l'indice d'inconsistance.

L'on constate toutefois que le système effectif choisi pour l'échantillon n_h /n correspond grandement à la part stratifiée N_h /N de l'ensemble (cf. tableau 7).

De la sorte, l'équivalence du sondage stratifié a priori et du sondage stratifié a posteriori était assurée.

Une distorsion de l'estimation de l'indice d'inconsistance ne provenait pas de cette déviation dans les conditions; comme l'on peut supposer qu'existe une faible corrélation positive, l'on peut éventuellement prévoir une légère sous-évaluation.

IV. Conclusion de la seconde partie

Rappelons les points essentiels de cette deuxième partie.

Pour calculer l'exactitude des données (les erreurs systématiques étant exclues) qui ont été collectées à l'aide d'enquêtes périodiques à l'occasion des Recensement agricoles en Belgique, il a fallu concevoir un modèle d'enquête répétée et l'appliquer par la suite. Celui-ci permettait d'estimer, à partir de deux enquêtes, les variabilités de réponses de valeurs de caractéristiques, constantes dans le temps (les superficies agricoles de chaque exploitation). Le Recensement agricole proprement dit, réparti en classes de grandeur (en d'autres termes l'enquête originale), faisait fonction d'enquête principale et la soi-disant enquête-pilote (un sondage), stratifiée selon les mêmes classes de grandeur, faisait fonction d'enquête répétée. En laissant tomber d'une part l'ajustement (constance dans le temps des valeurs des caractéristiques), l'on a pu simplifier le modèle par rapport à des applications antérieures présentant des caractéristiques variant dans le temps; d'autre part, l'application d'un sondage stratifié a nécessité l'adaptation du système des enquêtes répétées à ces données. A côté de cet ajustement formel, il a fallu, lors des estimations, prendre en considération l'influence de classifications erronées survenant éventuellement et en définir l'ordre de grandeur.

L'importance des variabilités estimées a semblé restreinte comme on pouvait s'y attendre dans le cas d'une caractéristique constante dans le temps.

Dans ce contexte, il faut encore préciser que des variabilités entre classes de grandeur sont d'autant plus grandes que la répartition en classes de grandeur est plus petite, et inversément. Pour cette raison, il faut si possible définir la répartition des classes de grandeur de telle sorte que la variabilité entre ces classes de grandeur ne perturbe pas la représentativité des résultats.

Tableau 3 : Recensement agricole et horticole du 15 mai 1985 en Belgique

Catégorie 1 (partiellement ajustée)

a) Variance des réponses (s^2_{Rh})

Classe de-grandeur SAU en ha	Code 260 Superifie agricole utilisée en a			Code 014 Superficie froment d'hiver en a			Code 018 Superficie orge d'hiver en a			Code 026 Superficie betteraves en a			Code 038 Superficie plantes industrielles en a		
h	$\hat{\bar{Y}}_h$	s_{Rh}	s_{Rh}(%)	$\hat{\bar{Y}}_h$	s_{Rh}	s_{Rh}(%)	$\hat{\bar{Y}}_h$	s_{Rh}	s_{Rh}(%)	$\hat{\bar{Y}}_h$	s_{Rh}	s_{Rh}(%)	$\hat{\bar{Y}}_h$	s_{Rh}	s_{Rh}(%)
0,01 - < 10	561,661	179,5907	32,0	34,297	51,3806	149,8	29,721	17,3137	58,3	17,441	52,8064	302,8	21,604	52,1473	241,1
10 - < 30	1733,476	277,0808	16,0	156,185	75,1507	48,1	100,809	35,5295	35,2	112,648	42,5978	37,8	123,635	44,0996	35,7
30 - < 100	4654,809	448,1179	9,6	753,232	150,8124	20,0	399,044	113,0523	28,3	531,369	166,0172	31,2	576,146	147,3442	25,6
100 - < +	13440,745	1119,7197	8,3	3992,731	529,5510	13,3	2014,649	507,0633	25,2	2639,306	426,5889	16,2	3121,976	404,8029	13,0
Total	2037,400	313,6269	15,4	270,198	103,7862	38,4	154,309	76,0328	49,3	187,588	94,0272	50,1	207,658	87,1311	41,9

SAU : superficie agricole utilisée

Moyenne $\hat{\bar{Y}}_h = \dfrac{\hat{Y}_{2h} + \hat{Y}_{1h}}{2 \cdot N_{h(85)}}$; $s_{Rh}(\%) = \dfrac{s_{Rh}}{\hat{\bar{Y}}_h}(\%)$; $\hat{\bar{Y}}_h$ moyenne commune issue des estimations des moyennes de l'enquête principale (1) et de l'enquête-pilote (2)

b) Variabilité des réponses

Classe de-grandeur SAU en ha	Code 260 Superifie agricole utilisée en a			Code 014 Superficie froment d'hiver en a			Code 018 Superficie orge d'hiver en a			Code 026 Superficie betteraves en a			Code 038 Superficie plantes industrielles en a		
h	$\hat{\bar{Y}}_h$	$\hat{V}^{1/2}_{Rh}$	$\hat{V}^{1/2}_{Rh}$(%)	$\hat{\bar{Y}}_h$	$\hat{V}^{1/2}_{Rh}$	$\hat{V}^{1/2}_{Rh}$(%)	$\hat{\bar{Y}}_h$	$\hat{V}^{1/2}_{Rh}$	$\hat{V}^{1/2}_{Rh}$(%)	$\hat{\bar{Y}}_h$	$\hat{V}^{1/2}_{Rh}$	$\hat{V}^{1/2}_{Rh}$(%)	$\hat{\bar{Y}}_h$	$\hat{V}^{1/2}_{Rh}$	$\hat{V}^{1/2}_{Rh}$(%)
0,01 - < 10	561,661	6,1190	1,1	34,297	1,7508	5,1	29,721	0,5885	2,0	17,441	1,8039	10,3	21,604	1,7813	8,2
10 - < 30	1733,476	6,7697	0,4	156,185	1,8472	1,2	100,809	0,8739	0,9	112,648	1,0469	0,9	123,635	1,0823	0,9
30 - < 100	4654,809	13,2038	0,3	753,232	4,4712	0,6	399,044	3,3494	0,8	531,369	4,9211	0,9	576,146	4,3667	0,8
100 - < +	13440,745	95,1607	0,7	3992,731	45,2001	1,1	2014,649	43,0279	2,1	2639,306	36,6873	1,4	3121,976	34,8116	1,1
Total	2037,400	4,6616	0,2	270,198	1,4288	0,5	154,309	0,9015	0,6	187,588	1,2596	0,7	207,658	1,1827	0,6

$\hat{V}^{1/2}_{Rh}(\%) = \dfrac{\hat{V}^{1/2}_{Rh}}{\hat{\bar{Y}}_h}(\%)$; $\hat{\bar{Y}}_h$ moyenne commune issue des estimations des moyennes de l'enquête principale (1) et de l'enquête-pilote (2)

Tableau 4 : Superficie cultivée de diverses cultures (Y_{1h}) selon le Recensement agricole et horticole du 15 mai 1985 en Belgique, de même que la variabilité des réponses ($\hat{V}_{Rh}^{1/2}(\%)$)

Catégorie 1

Classe de grandeur d'après la superficie agricole utilisée en ha	Nombre d'exploitation d'après le recensement du 15 mai 1985 (enquête principale)	Superficie agricole utilisée en a		Superficie froment d'hiver en a		Superficie orge d'hiver en a		Superficie betteraves en a		Supeficie plantes industrielles en a	
h	$N_{h\|85}$	Y_{1h} ha	$\hat{V}_{Rh}^{1/2}(\%)$	Y_{1h} ha	$\hat{V}_{Rh}^{1/2}(\%)$	Y_{1h} ha	$\hat{V}_{Rh}^{1/2}(\%)$	Y_{1h} ha	$\hat{V}_{Rh}^{1/2}(\%)$	Y_{1h} ha	$\hat{V}_{Rh}^{1/2}(\%)$
0,01 - < 10	19 284	105371,89	1,1	6170,83	5,1	5764,27	2,0	3131,92	10,3	3981,56	8,2
10 - < 30	29 844	538202,54	0,4	47559,13	1,2	30778,14	0,9	33995,47	0,9	38169,13	0,9
30 - < 100	11 558	537080,05	0,3	87662,56	0,6	46222,89	0,8	59096,98	0,9	65615,15	0,8
10 - < +	635	85176,63	0,7	24854,67	1,1	13497,32	2,1	17443,89	1,4	20437,65	1,1
Total	$N_{185} = 61321$	1265831,11	0,2	166247,19	0,5	96262,62	0,6	113668,26	0,7	128203,49	0,6

Remarque:

$N_{h|85}$ =données selon le recensement au 15 mai 1985.

Tableau 5 : Recensement agricole et horticole du 15 mai 1979 en Belgique

Catégorie 1

Définition de la variance des réponses et de la variabilité des réponses - Caractéristique : superficie froment d'hiver

Table de calcul 5a : Valeurs déclarées individuelles pour la classe de grandeur h = 0,01 ha - < 10 hade superficie agricole utilisée - extrait.

(Dans cette calsse de grandeur beaucoup d'exploitations ne cultivent pas de froment, d'où peu de données).

Numéro courant de l'exploitation (i)	Code de l'exploitation (i) associé au code communal / Code de l'exploitation	Code 014 Superficie froment d'hiver — Valeur déclarée en a : 04 mai y_{i2AFh}	15 mai y_{i1AFh}	$d_{iAFh} = y_{i1AFh} - y_{i2AFh}$	Variance des réponses individuelle $s^2_{R_iAFh} = \frac{d^2_{iAFh}}{2}$	(i)	Code de l'exploitation (i)	Code 014 superficie froment d'hiver — 04 mai y_{i2NFh}	15 mai y_{i1NFh}	d_{iNFh}	$s^2_{R_iNFh}$
		- Ancien fichier -						- Nouveaux venus -			
	1 Province d'Anvers						1 Province d'Anvers				
1	11001 Aartselaar/0004					1	11004 Boechout/0259				
2	11009 Brecht/0176					2	11009 Brecht/0455				
3	11016 Essen/0007					3	11057 Malle/0196				
4	11022 Kalmthout/0023					4	12014 Heist-op-den-Berg/0731				
5	11035 Ranst/0050					5	12014 Heist-op-den-Berg/0741				
6	11037 Rumst/0028					6	13002 Baarle-Hertog/0136				
7	11053 Wuustwezel/0191					7	13003 Balen/0073				
8	11055 Zoersel/0064					.					
9	12002 Berlaar/0173					.					
10	12014 Heist-op-den-Berg/0335					.					
11	12025 Malines/0048	210	210	0	0,0	.					
12	12025 Malines/0390	50	40	-10	50,0	.					
13	12026 Nijlen/0142					.					
14	12034 Sint-Amands/0183										
15	12035 Sint-Katelijne-Waver/0101										
.	9 Province de Namur						9 Province de Namur				
412	91054 Gedinne/0060					215	91054 Gedinne/0219				
413	91064 Havelange/0222					216	91059 Hamois/0231				
414	92035 Eghezee/0215	200	150	-50	1250,0	217	91064 Havelange/0261				
415	92045 Floreffe/0072	40	40	0	0,0	218	91064 Havelange/0266				
416	92094 Namur/0019	700	700	0	0,0	219	91114 Rochefort/0344				
417	92137 Sambreville/0005					220	92087 Mettet/0340				
418	92137 Sambreville/0082					221	92138 Fernelmont/0184	200	115	-85	3612,5
419	92141 La Bruyère/0009					222	93088 Walcourt/0372				
420	92142 Gembloux/0012										
421	93010 Cerfontaine/0209										
422	93056 Philippeville/0005										
423	93088 Walcourt/0139										
	Total	Σy_{i2AFl} = 15 292	Σy_{i1AFl} = 13 804	Σd_{iAFl} = -1488	$\Sigma s^2_{R_iAFl}$ = 1 126 658,0		Total	Σy_{i2NFl} = 5 967	Σy_{i1NFl} = 5 612	Σd_{iNFl} = 619	$\Sigma s^2_{R_iNFl}$ = 138 624,5

Le classement des exploitations en classes de grandeur résulte de la valeur déclarée de la superficie agricole utilisée dans l'enquête principale du 15 mai 1985.

Table de calcul 5b : Valeurs d'appui pour calculer la variance des réponses s^2_{Rh} dans la classe $h = 1$ et calcul de celle-ci.

-Classe de grandeur 0,01 ha - < 10 ha de superficie agricole utilisée -
(chiffre partiellement arrondi)

Code 014 Superficie froment d'hiver

Ancien fichier	Valeur déclarée en a : 04 mai	Valeur déclarée en a : 15 mai	Variance des réponses simple
$N_{AFh} = N_{AF1}$ resp. $N_{AF1(85)}$ = 19.017 exploitations (chiffre calculé) $n_{AFh} = n_{AF1} = 423$ exploitations sondées (chiffre effectif) Facteur d'extrapolation de l'ancien fichier calculé $\frac{N_{AF1(85)}}{n_{AF1}} = \frac{19.017}{423} = 44{,}95744681$ Nouveaux venus $N_{NFh} = N_{NF1}$ resp.= $N_{NF1(85)} = 222$ exploitations (chiffre effectif) - Pas d'échantillon - Recensement -	$\hat{Y}_{AF21} = \frac{N_{AF1(85)}}{n_{AF1}} \sum_{i=1}^{n_{AF1}} y_{i2AF1}$ = 687.489,277 $Y_{NF21} = \sum_{i=1}^{N_{NF1(85)}} y_{i2NF1}$ = 5.967	$\hat{Y}_{AF11} = \frac{N_{AF1(85)}}{n_{AF1}} \sum_{i=1}^{n_{AF1}} y_{i1AF1}$ = 620.592,596 $Y_{NF11} = \sum_{i=1}^{N_{NF1(85)}} y_{i1NF1}$ = 5.612	$s^2_{RAF1} = \frac{1}{n_{AF1}} \sum_{i=1}^{n_{AF1}} s^2_{RiAF1}$ = 2.663,49409 $s^2_{RNF1} = \frac{1}{N_{NF1(85)}} \sum_{i=1}^{N_{NF1(85)}} s^2_{RiNF1}$ = 624,43468
TOTAL $N_{h(85)} = N_{AF1(85)} + N_{NF1(85)}$ = 19.017 + 222 = 19.239 exploitations (chiffre effectif) (Ancien fichier / Nouveau fichier) $n_1 = n_{AF1} + N_{NF1(85)}$ = 423 + 222 = 645 exploitations (chiffre effectif) $N_{1\|85}$ = nombre d'exploitations d'après le recensement du 15 mai 1985 (enquête principale) dans la classe de grandeur $h = 1$ = 19.284 exploitations	Valeur estimée de la valeur totale en a $\hat{Y}_{21} = \hat{Y}_{AF21} + Y_{NF21}$ = 693.456,277 Valeur moyenne totale en a $\hat{\bar{Y}}_1 = \frac{\hat{Y}_{21} + \hat{Y}_{11}}{2.N_{1(85)}} = 34{,}2965$ (valeur de l'enquête principale au 15 mai 1985 : Y_{11} = 617.083 a - recensement)	$\hat{Y}_{11} = \hat{Y}_{AF11} + Y_{NF11}$ = 626.204,596	Variance des réponses de la classe de grandeur 1 $s^2_{R1} = \frac{1}{N_{1(85)}} \left[N_{AF1(85)} . s^2_{RAF1} + N_{NF1(85)} . s^2_{RNF1} \right]$ = 2.639,96210 $s_{R1} = 51{,}38056$ Variance des réponses relative (%) $s_{R1}(\%) = \frac{s_{R1}}{\hat{\bar{Y}}_1} = 1{,}49813 \approx 150\%$

Remarque : l'ancien fichier (chiffre calculé) = résultat du recensement au 15 mai 1985 (19.284 exploitations) - nouveaux venus (chiffre réel) du 15 mai 1983 au 15 mai 1985 (267 exploitations) = 19.017 exploitations.

Pour la signification d' "effectif", voir la remarque dans le tableau 5c.

Tableau 5c : Calcul de la variabilité des réponses (variance d'échantillonnage des réponses) $\hat{V}_{Rh}$

$$\hat{V}_{Rh} \approx \frac{1}{N^2_{h(85)}}\left[N^2_{AFh(85)} \cdot \frac{s^2_{RAFh}}{2n_{AFh}} + N^2_{NFh(85)} \cdot \frac{s^2_{RNFh}}{2N_{NFh(85)}}\right]$$

(formule partiellement arrondie)

classe de grandeur 0,01 ha - < 10 ha SAU($h = 1$)

Ancien fichier $N_{AFh} = N_{AF1}$ resp. $N_{AF1(85)} = 19.017$ exploitations (chiffre calculé) $n_{OBh} = n_{OB1} = 423$ exploitations sondées (chiffre effectif) **Nouveaux venus** $N_{NFh} = N_{NF1}$ resp. $N_{NF1(85)} = 222$ exploitations (chiffre calculéf) Pas de sondage - Recensement	Code 014 Superficie froment d'hiver $S^2_{RAFB1} = 2.663,49409$ $N^2_{AF1(85)} \cdot \frac{s^2_{RAFB1}}{2n_{AF1}} = 1.134.561.547$ $s^2_{RNF1} = 624,43468$ $N^2_{NF1(85)} \cdot \frac{s^2_{RNF1}}{2N_{NF1(85)}} = 69.312,25$
TOTAL $N_{h(85)} = N_{1(85)} = N_{AF1(85)} + N_{NF1(85)} = 19.239$ exploitations (chiffre effectif) Remarque : "effectif" signifie : que tant dans l'enquête principale (1) que dans l'enquête-pilote (2) il y a des réponses realtives à la caractéristique superficie agricole utilisée	$[\ldots] = 1.134.630.859,25$ $\hat{V}_{Rh} = \frac{1}{N^2_{1(85)}}[\ldots] = 3,0654173$ $\hat{V}^{1/2}_{R1} \approx 1,75083$ $\hat{V}^{1/2}_{R1}(\%) \approx \frac{\hat{V}^{1/2}_{R1}}{\hat{\bar{Y}}_1} = 0,05105 \approx$ 5,1 %

Table de calcul 5d : Calcul des valeurs estimées pour la valeur totale de toutes les classes de grandeur $L = 4$

Valeur totale des exploitations $N_{(85)} = \sum_{h=1}^{L=4} N_{h(85)} = N_{1(85)} + N_{2(85)} + N_{3(85)} + N_{4(85)}$; valeur estimée de la moyenne totale

$$\hat{Y}_1 = \sum_{h=1}^{L=4} \hat{Y}_{1h};\ \hat{Y}_2 = \sum_{h=1}^{L=4} \hat{Y}_{2h};\qquad \hat{\bar{Y}} = \frac{\hat{Y}_2 + \hat{Y}_1}{2N_{(85)}};$$

(Chiffre partiellement arrondi)

Classes de grandeur d'après la superficie agricole utilisée en ha h	Nombre d'exploitations selon le recensement au 15 mai 1985 (enquête principale) $N_{h\|85}$	Nombre d'exploitations d'après l'ancien fichier nouveaux venus ayant répondu (chiffre effectif) $N_{h\,(85)}$	Code 014 Superficie froment d'hiver - valeur déclarée en a (valeur estimée) $\hat{Y}_{2h}$	 $\hat{Y}_{1h}$
0,01 - < 10	19 284	19 239	693 456,28	626 204,60
10 - < 30	29 844	29 814	4 703 162,92	4 609 806,37
30 - < 100	11 558	11 547	8 812 072,18	8 583 064,66
100 - < +	635	634	2 512 779,67	2 550 002,42
Total	$N_{\|85}$ = 61 321	$N_{(85)}$ = 61 234	$\hat{Y}_2$ = 16 721 471,05	$\hat{Y}_1$ = 16 369 078,04

Remarque :
$N_{h|85}$ = résultat du recensement au 15 mai 1985
$N_{h(85)}$ = ancien fichier + nouveaux venus (chiffre effectif)

Y_1 = 16 624 719 a
= 166 247,19 ha
Valeur moyenne totale
$\hat{\bar{Y}} = \frac{\hat{Y}_2 + \hat{Y}_1}{2N_{(85)}}$ = 270,20 a

Table auxiliaire de calcul

Nombre d'exploitations (ancien et nouveau fichier) et nombre d'exploitations de l'échantillon selon les classes de grandeur

Classe de grandeur (SAU) en ha h	Nombre d'exploitations au 15 mai 1985 d'après le recensement $N_{h\|85}$ (1)	Nouveaux venus dans la période du 15.05.1983 au 15.05.1985 chiffre réel (2a)	 chiffre effectif (2b) (N_{NFh})	Ancien fichier (chiffre calculé) au 15 mai 1985 (3) = (1) - (2a).(N_{AFh})	Nombre d'exploitations de l'échantillon Cadre ancien fichier chiffre réel	 chiffre effectif (n_{OBh})
0,01 - < 10	19 284	267	222	19 017	513	423
10 - < 30	29 844	222	192	29 622	949	820
30 - < 100	11 558	69	58	11 489	651	563
100 - < +	635	6	5	629	70	67
Total	61 321	564	477	60 757	2 183	1 873

Remarque : Le nombre d'exploitaitons d'après le recensement du 15 mai 1985 et le nombre de nouveaux venus ne sont pas exacts et font apparaître des erreurs d'enregistrement; l'expérience nous a toutefois appris qu'elles sont minimes. "Réel" signifie que l'exploitation existait dans la période du 15.05.1983 au 15.05.1985 ou a été ajoutée en tant que "nouveau venu" durant cette période.

Table de calcul 5e : Calcul de la variance des réponses et de la variabilité des réponses (variance d'échantillonnage des réponses) pour toutes les classes de grandeur L = 4

Variance des réponses : $s_R^2 = \frac{1}{N_{(85)}} \sum_{h=1}^{L=4} N_{h(85)} \cdot s_{Rh}^2$

Variabilité des réponses : $\hat{V}_R \approx \frac{1}{N_{(85)}^2} \sum_{h=1}^{L=4} N_{h(85)}^2 \cdot \frac{1}{N_{h(85)}^2} \underbrace{\left[N_{AFh(85)}^2 \cdot \frac{s_{RAFh}^2}{2n_{AFh}} + N_{NFh(85)}^2 \cdot \frac{s_{RNFh}^2}{2N_{NFh(85)}} \right]}_{\hat{V}_{Rh}}$

(formule partiellement arrondie)

Classe de grandeur selon la superficie agricole utilisée en ha h	Nombre d'exploitations calculé selon l'ancien fichier et des nouveaux venus ayant répondu (chiffre effectif) $N_{h(85)}$	Code 014 Superficie froment d'hiver s_{Rh}^2	 $N_{h(85)}^2 \cdot s_{Rh}^2$	Code 014 Superficie froment d'hiver $\hat{V}_{Rh}$	 $N_{h(85)}^2 \cdot \hat{V}_{Rh}$
0,01 - < 10	19 239	2 639,962115	50 790 230,95	3,06542	1 134 630 859
10 - < 30	29 814	5 647,62847	168 378 395,2	3,41210	3 032 930 726
30 - < 100	11 547	22 744,37668	262 629 317,5	19,99168	2 665 555 336
100 - < +	634	280 424,21560	177 788 952,7	2 039,48627	819 783 744
Total	$N_{(85)} = 61\,234$	----------	659 586 896,35	----------	7 652 900 665

$s_R^2 = \frac{1}{N_{(85)}} \sum_{h=1}^{L=4} N_{h(85)} s_{Rh}^2 = 10\,771{,}57946$

$s_R = 103{,}78622$

$s_R(\%) = \frac{s_R}{\hat{\bar{Y}}} = \frac{103{,}78622}{270{,}20} = 0{,}38411 \approx$ 38,4%

$\hat{V}_R \approx \frac{1}{N_{(85)}^2} \sum_{h=1}^{L=4} N_{h(85)}^2 \hat{V}_{Rh} = 2{,}0409$

$\hat{V}_R^{1/2}(\text{Racine}) = 1{,}42876$

$\hat{V}_R^{1/2}(\%)(\text{Racine}) = \frac{\hat{V}_R^{1/2}}{\hat{\bar{Y}}} = \frac{1{,}42876}{270{,}20} = 0{,}00529 \approx$ 0,5%

Tableau 6 : Recensement agricole et horticole du 15 mai 1985 en Belgique (catégorie 1)

Tableau a

Table de contingence pour la répartition des anciennes et nouvelles exploitations selon la valeur déclarée de la superficie agricole utilisée lors de l'enquête principale (15 mai 1985) et lors de l'enquête-pilote (4 mai 1985) dans les classes de grandeur superficie agricole utlisée

	Ancien fichier (AF) (sondage)		Nouveaux venus (NF) (recensement)		
Enquête-pilote classes de grandeur superficie agricole utilisée ha h / Enquête principale classes de grandeur superficie agricole utilisée ha h	0,01 - < 10	10 - < 30	30 - < 100	100 - < +	Σ
0,01 - < 10	410 *213*	36 *7*	0 *2*	0 *0*	446 *222*
10 - < 30	12 *9*	767 *173*	27 *8*	0 *0*	806 *190*
30 - < 100	1 *0*	17 *12*	533 *47*	8 *1*	559 *60*
100 - < +	0 *0*	0 *0*	3 *1*	*59* *4*	62 *5*
Σ	423 *222*	820 *192*	563 *58*	*67* *5*	1.873 *477*

Les chffres relatifs aux "nouveaux venus" sont en italique.

Tableau b

Indice d'inconsistance

- Index of Inconsistency -

Classe de grandeur superficie agricole utilisée ha	N_{AFh}	N_{NFh}	N_h $(= N_{AFh} + N_{NFh})$	$\hat{I}_{AFh}$	I_{NFh}	$\hat{I}_{gh}$	$\hat{\bar{I}}_h$
0,01 - < 10	19 017	222	19 239	7,34	7,58	7,35	7,35
10 - < 30	29 622	192	29 814	10,00	15,72	10,03	10,04
30 - < 100	11 489	58	11 547	7,13	23,21	7,17	7,21
100 - < +	629	5	634	8,83	20,21	8,89	8,92
Total	60 757	477	$N = 61\,234$			$\hat{I}_g = 8{,}53$	$\hat{\bar{I}} = 8{,}63$

$\hat{I}_{AFh}$ resp. I_{NFh} = indice d'inconsistance (%) selon l'ancien resp. le nouveau fichier dans la classe de grandeur h

$\hat{I}_{gh}$ resp. $\hat{I}_g$ = moyennes pondérées géométriquement (%), $\hat{\bar{I}}_h$ resp. $\hat{\bar{I}}$ = moyenne arithmétique pondérée

$$\log \hat{I}_{gh} = \frac{N_{AFh} \log \hat{I}_{AFh} + N_{NFh} \log I_{NFh}}{N_h}, \hat{\bar{I}}_h = \frac{N_{AFh} \hat{I}_{AFh} + N_{NFh} I_{NFh}}{N_h}$$

$$\log \hat{I} = \frac{\sum_{h=1}^{L=4} N_h \log \hat{I}_{gh}}{N}, \hat{\bar{I}} = \frac{\sum_{h=1}^{L=4} N_h \hat{\bar{I}}_h}{N}$$

Tableau 7 : Recensement agricole et horticole au 15 mai 1985 en Belgique (catégorie 1) Nombre d'exploitations N_h et nombre d'exploitations de l'échantillon n_h d'après les classes de grandeur superficie agricole utilisée (SAU)

Classe de grandeur SAU en ha h	N_{AFh}	N_{NFh}	N_h	n_{AFh}	n_h	$\frac{N_h}{N}$	$\frac{n_h}{n}$
0,01 - < 10	19 017	222	19 239	423	645	0,31	0,27
10 - < 30	29 622	192	29 814	820	1 012	0,48	0,43
30 - < 100	11 489	58	11 547	563	621	0,19	0,26
100 - < +	629	5	634	67	72	0,01	0,03
Σ	60 757	477	61 234	1 873	2 350	1	1

Etant donné : $N_h = N_{AFh} + N_{NFh}$; $n_h = n_{AFh} + N_{NFh}$

Tableau 8a : Recensement agricole et horticole au 15 mai 1985 en Belgique (catégorie 5)

Variance des réponses (s^2_{Rh}) et variabilité des réponses $(\hat{V}_{Rh})$ (chiffres partiellement arrondis)

Classe de grandeur superficie agricole utilisée en ha h	Variance des réponses						Variabilité des réponses			
	Code 260 Superficie agricole utilisée en a			Code 014 Superficie froment d'hiver en a			Code 260 Superficie agricole utilisée		Code 014 Superficie froment d'hiver	
	$\overline{Y}_h$	s_{Rh}	$s_{Rh}(\%)$	$\overline{Y}_h$	s_{Rh}	$s_{Rh}(\%)$	$\hat{V}^{\frac{1}{2}}_{Rh}$	$\hat{V}^{\frac{1}{2}}_{Rh}(\%)$	$\hat{V}^{\frac{1}{2}}_{Rh}$	$\hat{V}^{\frac{1}{2}}_{Rh}(\%)$
0,01 - < 10	389,925	703,112	180,3	41,553	14,391	34,6	30,484	7,8	0,624	1,5
10 - < 30	1 852,695	2 113,529	114,1	290,943	90,449	31,1	160,226	8,6	6,857	2,3
30 - < 100	4 618,146	272,757	5,9	834,458	206,551	24,8	39,369	0,9	29,813	3,6
100 - < +	164,400	0	0	45,000	0	0	0	0,0	0	0,0
Total	1 037,513	1 175,045	113,3	162,414	68,829	42,4	42,736	4,1	2,503	1,5

$\overline{Y}_h = \frac{Y_{1h} + Y_{2h}}{2N_h}$ = **valeur moyenne commune provenant de l'enquête principale** (Y_{1h}) **et de l'enquête pilote** (Y_{2h}) **(par exploitation)**

Abréviations des mesures de superficie : ha = hectare = 100 are, a = are = 100 m²

Tableau 8b : Surface cultivée de diverses cultures d'après le recensement agricole et horticole en Belgique au 15 mai 1985 et l'enquête-pilote du 4 mai 1985 (Y_{2h}) ainsi que les variabilités des réponses $(\hat{V}^{\frac{1}{2}}_{Rh}(\%))$.

Catégorie 5

Classe de grandeur superficie agricole utilisée en ha h	Nombre effectif d'exploitation au 15 mai 1985 $N_h(85)$	Code 260 superficie agricole utilisée Valeur totale des réponses effectives (déclarations) en ha			Code 014 superficie froment d'hiver Valeur totale des réponses effectives (déclarations) en ha		
		15 mai Y_{1h}	04 mai Y_{2h}	$\hat{V}^{\frac{1}{2}}_{Rh}(\%)$	15 mai Y_{1h}	04 mai Y_{2h}	$\hat{V}^{\frac{1}{2}}_{Rh}(\%)$
0,01 - < 10	266 (314)	944,10 (1 159,48)	1 130,30	7,8	111,20 (135,79)	109,87	1,5
10 - < 30	87 (105)	1 477,20 (1 717,22)	1 746,49	8,6	256,21 (315,14)	250,03	2,3
30 - < 100	24 (21)	1 099,62 (849,72)	1 117,09	0,9	206,99 (193,09)	193,55	3,6
100 - < +	1 (1)	164,40 (106,28)	164,40	0	45,00 (33,90)	45,00	0
Total	$N_{(85)} = 378$ $(N_{85} = 441)$	3 685,32 (3 832,70)	4 158,28	4,1	619,40 (677,92)	608,45	1,5

(...) Valeur totale des résultats de comptage} selon le recensement agricole et horticole du 15 mai 1985

(...) = N_{h85} nombre d'exploitations. }

Comme l'enquête-pilote était volontaire, tous les chefs d'exploitation contactés n'ont pas renvoyé à l'I.N.S. le questionnaire de l'enquête-pilote; il en résulte de manière presque absolue que $N_{h(85)} < N_{h85}$ et que la valeur déclarée totale Y_{1h} < (....) les résultats du recensement au 15 mai 1985. Par "effectif"l'on entend que, dans le calcul des valeurs totales, moyennes et de variance, on a seulement repris les exploitations pour lesquelles apparaît tant dans l'enquête principale que dans l'enquête-pilote au minimum la caractéristique superficie agricole utilisée.

CONCLUSION GENERALE

Dans les deux parties de ce texte, l'on a exposé les fondements de l'approche théorique et de l'application pratique des modèles d'enquête avec répétition, avec ou sans ajustement et avec ou sans stratification. L'on a démontré - concernant des conditions d'enquête diverses et des caractéristiques distinctes - l'efficacité de cette manière de procéder dans l'estimation de la variabilité des réponses mesurant l'influence des erreurs de déclaration. Ces modèles ont été conçus et appliqués avec succès en Belgique dans le cadre des Recensements agricoles et horticoles des 15 mai 1979 et 15 mai 1985. L'on a présenté, examiné et évalué les principaux résultats de ces recensements.

Dans la mesure du possible, il est recommandé d'évaluer de temps en temps, selon une forme méthodique analogue, basée sur le genre de modèles présentés ici, la qualité statistique de données provenant d'enquêtes statistiques périodiques. De la sorte, l'on dispose d'une méthode pour contrôler qualitativement des données statistiques et informer de manière satisfaisante non seulement les nombreux utilisateurs de la statistique officielle mais aussi les producteurs de données.

Nous rappelons que ces travaux sont le fruit d'une féconde et stimulante collaboration entre des chercheurs de l'Université de Tubingue et des fonctionnaires de l'Institut national de statistique à Bruxelles.

Des recherches communes comme celles-ci sont une illustration exemplaire de coopération européenne.

Résumé

Dans la méthodologie des enquêtes statistiques la variabilité des réponses des déclarations individuelles n'est pas quantité négligeable; on l'estime au moyen d'enquêtes répétées . Comme en Europe occidentale beaucoup de résultats statistiques s'obtiennent en faisant remplir les questionnaires par les enquêtés, nous n'avons dans cet article traité que des modèles d'enquête sans intervention de recenseurs et d'intervieweurs.

Dans ce concept, lorsqu'on établit les échantillons, il y a deux sortes d'enquêtes répétées, des enquêtes ayant lieu le même jour ou des enquêtes se déroulant sur plusieurs jours. Pour des raisons pratiques l'on est fréquemment obligé dans le cas de caractéristiques non-constantes de n'utiliser que des modèles à plusieurs jours de référence. Dans ces cas il faut ramener les valeurs déclarées des divers recensements (répétés) à un même jour d'enquête. Dans la deuxième partie, l'on a exposé la méthodologie de différentes enquêtes répétées, ainsi qu'à quelles conditions et de quelle façon l'on peut effectuer un ajustement (précis ou seulement sous forme d'estimation) à l'intérieur du modèle. A l'aide de deux exemples pratiques tirés de la statistique agricole belge (au 15 mai 1979 et au 15 mai 1985), l'on a exposé les applications correspondantes du modèle et communiqué les résultats.

Dans la deuxième partie de cette étude, l'on a présenté des recherches sur l'estimation de la variabilité des réponses en cas de valeurs de caractéristiques constantes dans le temps. Les caractéristiques en question étaient les diverses superficies cultivées par exploitation. A côté de l'enquête principale (le Recensement agricole et horticole belge au 15 mai 1985), l'on a effectué séparément comme enquête répétée un sondage stratifié économico-administratif. Les résultats classés par ordre de grandeur ont aussi été examinés pour voir si les erreurs de déclaration avaient entrainé des classifications erronées. Sur la base des valeurs enquêtées l'on a pu définir des estimations plausibles pour les variances et les variabilités des réponses. Sans problèmes l'on peut en modifiant (légèrement) le modèle développer des concepts analogues lors de la définition de grandeurs de mouvements; ce sujet a été brièvement abordé dans la partie I.

Les parties I et II ensemble donnent une indication sur les possibilités de mise en pratique de cette forme de contrôle de qualité de données statistiques au moyen d'enquêtes réalisées de manière continue.

BIBLIOGRAPHIE

ANDERSON SEN., 0. (1957) : Probleme der Statistischen Methodenlehre in den Sozialwissenschaften, Einzelschriften der Deutschen Statistischen Gesellschaft, Nr. 6, Physica Würzburg (3. Aufl.), pp. 3-14, 80-95, 227-235 et 244-249.

BAILAR, B.A. (1968) : Recent Research in Reinterview Procedures, Journal of the Amercian Statistical Association, Vol. 63, pp. 41-63.

BAILAR, B.A., DALENIUS, T. (1969) : Estimating the Response Variance Components of U.S. Bureau of the Census Survey Model, Sankya, Vol. 31, Series B, pp. 341-360.

BAMBERG, G., BAUR, F. (1989) : Statistik, München, Wien (6. Aufl.), pp. 1-10.

BECKMANN, M., WIEGERT, R. (1987) : Statistische Erhebungen : Methoden und Ergebenisse - Ausgewählte Schriften, von Heinrich Strecker Schriftenreihe : Angewandte Statistik und Ökonometrie, Heft 30, Vandenhoeck u. Ruprecht, Göttingen, pp. 83-86 et 131-211.

BERSHARD, M. (1969) : The Index of Inconsistency for an L-Fold Classification System, L ≥ 2 U.S. Bureau of the Census, Technical Notes 2, pp. 1-3.

BIEMER, P.P., GROVES, R.M., LYBERG, L.E., MATHIOWETZ, N.A., SUDMAN, S. (1991) : Measurement Errors in Surveys. John Wiley, New York, Toronto, Singapore.

CHAUDHURI, A., MUKERJEE, R. (1988) : Randomized Response - Theory and Techniques. Marcel Dekker, New York, Basel.

COCHRAN, W.G. (1977) : Sampling Techniques (third edition), John Wiley, New York, pp. 127-131 (optimum stratification) et 134-135 (a posteriori stratification).

DALENIUS, T., GURNEY, M. (1951) : The problem of optimum stratification, II. Skandinavisk Akturietidskrift, Vol. 34, pp. 133-148.

DALENIUS, T. (1957) : Sampling in Sweden, Contributions to the methods and theories of sample survey practice, Almqvist and Wicksell, Stockholm, pp.159-186.

DALENIUS, T., HODGES, L. Jr. (1959) : Minimum variance stratification, Journal of the American Statistical Association, Vol. 54, pp. 88-101.

DALENIUS, T. (1985) : Elements of Survey Sampling, Notes prepared for the Swedish Agency for Research Cooperation with Developing Countries (SAREC), Stockholm, XIX Non-Response Errors ((XIX-1-XIX-9), XX Measurement Errors (XX-1-XX-9), II Bases for Planning a Sample Survey (II-1-II-5), III Criteria for Survey Sampling (III-1-III-3), IV Preparing the Sampling Frame (IV-1-IV-27).

DEININGER, R., SZAMEITAT, K. (1967) : Some Remarks on the Problem of Errors in Statistical Results. Proceedings of the 36th Session, Bulletin of the International Statistical Institute, Sydney, Vol. XLII, Book 1, pp.16-71.

DEMING, W.E. (1953) : On a probability mechanism to attain an economic balance between the resulting error of response and bias of non-response. Journal of the American Statistical Association, Vol. 48, pp. 743-772.

DESABIE, J. (1966) : Théorie et pratique des sondages, Dunod, Paris, pp. 134-172 et 399-468.

FELLEGI, I.P. (1964) : Response Variance and its Estimation, Journal of the Amercian Statistical Association, Vol. 59, pp.1016-1041.

FELLEGI, I.P. (1974) : An Improved Method of Estimating the Correlated Response Variance, Journal of the American Statistical Association Vol. 69, pp. 496-501.

FORSMAN, G. (1989) : Early Survey Models and their Use in Survey Quality Work, Journal of Official Statistics (JOS), Vol. 5, No. 1, Statistics Sweden, pp. 41-55.

FORSMAN, G. (1991) : Survey Models - a Review and some Applications to Reinterview Data, Department of Statistics, University of Lund (thèse de doctorat).

FORSMAN, G. (1993) : Recent Advances in Survey Error Modelling, Jahrbücher für Nationalökonomie und Statistik. Band 211-331-350. - Cet article contient une bibliographie de toutes les publications importantes en anglais en ce domaine.

GROBRAS, J.K. (1987) : Méthodes statistiques des sondages, Economica, Paris, pp. 113-127.

GROVES, R. (1989) : Survey Errors and Survey Costs, John Wiley, New York, pp. 324-329.

HANSEN, M., HURWITZ, W. (1946) : The Problem of Non-Response in Sample Surveys, Journal of the American Statistical Association, Vol. 41, pp. 517-529.

HANSEN, M., HURWITZ, W. et MADOW, M. (1953) : Sampling Survey Methods - Vol. I : Methods and Applications, Vol. II Theory, John Wiley, New York and London.

HANSEN, M., HURWITZ, W. (1961) : Measurement Errors in Censuses and Surveys, 32th Session of the International Statistical Institute, Bulletin of the International Statistical Institute, Tokyo, Vol. 38, pp. 359-374.

HANSEN, M., HURWITZ, W., JABINE, T. (1963) : The use of imperfect lists for probability sampling at the U.S. Bureau of the Census, Proceedings of the 34th Session of the International Statistical Institute, Bulletin of the International Statistical Institute, Ottawa, Vol. XL, Book I, pp. 497-517.

HANSEN, M., HURWITZ, W., PRITZKER, L. (1964) : The Estimation and Interpretation of Gross Differences and the Simple Response Variance, Contributions to Statistics, ed. C.R. Rao, Pergamon Press, Oxford, Calcutta, pp. 111-136.

HENDRICKS, W.A., (1949) : Adjustment for bias by non-response in mailed surveys, Agricultural Economics Research, Vol. 1, pp. 52-56.

HENDRICKS, W.A. (1956) : The Mathematical Theory of Sampling, The Scanecrow Press, New Brunswick N.J., pp. 339-356.

HOCHBERG, Y. (1977) : On the Use of Double Sampling Schemes in Analyzing Categorial Data with Misclassification Errors, Journal of the American Association, Vol. 72, pp. 914-921.

KANTOROWITZ, M. (1969) : Evaluation of the Census Data, State of Israel - Central Bureau of Statistics, Publication Nr. 40, Vol. I, pp. 93-112.

KISH, L. (1965) : Survey Sampling, John Wiley, New York, pp. 527-532.

KRÓTKI, K.P. (1978) : Estimation of Correlated Response Variance, Proceedings of Section on Survey Research Methods (American Statistical Association), pp. 609-614.

KRÓTKI, K.P., MACLEOD, A. (1979) : Two Methods of Measuring correlated Response Variance, American Statistical Association, Washington, Proceedings of the Section on Survey Research Methods, pp. 223-230.

KRUG, W., NOURNEY, M., SCHMIDT, J. (1994) : Wirtschafts- und Sozialstatistik : Gewinnung von Daten, Oldenbourg, München und Wien, 3. Auflage, pp. 187-200.

LESSLER, J.T., KALSBEEK, W.D. (1992) : Nonsampling Error in Surveys ; John Wiley, New York, Toronto, Singapore.

MADOW, W.G., NISSELSON, H., OLKIN, L. (1983) : Incomplete Data in Sample Surveys - Vol. 1 Report and case studies, Vol. 2 Theory and Bibliographies, Vol. 3 Proceedings of the Symposium, Academic Press, New York, London, Paris, Tokyo.

MAHALANOBIS, P.C. (1946) : Recent experiments in statistical sampling in the Indian Statistical Institute, Journal of Royal Statistical Society, Vol. 109, pp. 325-370.

O' MUIRCHEARTAIGH, C.A., (ed.) (1977) : The analysis of survey data, Vol. 2, John Wiley, New York, Brisbane, Toronto, Chapter 7, Response error, pp. 193-239.

PFANZAGL, J. (1968) : Theory of Measurement, Physica, Würzburg und Wien, 2. Auflage, pp; 191-214.

POKROPP, F. (1996) : Stichproben : Theorie und Verfahren, Oldenbourg, München und Wien, 2. Auflage, pp. 191-214 .

POLITZ, A., SIMMONS,W. (1949) : An attempt to get the "not-at-homes" into the sample without callbacks, Journal of the American Statistical Association, Vol. 44, pp. 9-31.

POLITZ, A., SIMMONS, W. (1950) : Note on attempt to get "not-at-homes" without callbacks, Journal of the American Statistical Association, Vol. 45, pp. 136-137.

PRESS, J.S. (1968) : Estimating from Misclassified Data, Journal of American Association, Vol. 63, pp. 123-133.

PRITZKER, L., HANSEN, R. (1962) : Measurement Errors in the 1960 Census of Population, Proceedings of the Social Statistics Section of the American Statistical Association, pp. 80-90.

RADERMACHER, W., (1990) : Das statistische Informationssystem zur Bodennutzung STABIS der amtlichen Statistik, Geo-Informations-Systemen, Heft 4, pp. 9-15.

RADERMACHER W, (1992a) : Methoden und Möglichkeiten der Qualitätsbeurteilung von statistischen Informationen aus der Fernerkundung, Jahrbücher für Nationalökonomie und Statistik, Band 209, pp. 169-179.

RADERMACHER, W. u.a. (1992b) : Neue Wege der raumbezogenen Statistik, Bd. 20 der Schrifenreihe Forum der Bundesstatistik, Hrsg. Statistisches Bundesamt, Wiesbaden.

RINNE, H. (1970) : Kontrollen der Ergebnisse von Volkszählungen und Volkswirtschaftlichen Gesamtrechnungen, Allgemeines Statistisches Archiv, Band 54, Heft 1, pp. 29-54.

SCHÄFFER, K.A., SZAMEITAT, K. (1963) : Imperfect frames in statistics and the consequences for their use in sampling. Proceedings of the 34th Session, Bulletin of the International Statistical Institute, Ottawa, Vol. XL, Book 1, pp. 517-538.

SCHNEEBERGER, H. (1985) : Maxima, Minima und Sattelpunkte bei optimaler Schichtung und optimaler Aufteilung, Allgemeines Statistisches Archiv, Band 69, Heft 3, pp. 286-297.

SCHNEEBERGER, H. (1991) : Some Comments on Sampling Optimization, Jahrbücher für National-ökonomie und Statistik, Band 208, pp. 67-80.

SCHNEEWEISS, H. (1990) : Ökonometrie, 4. Auflage, Physica, Heidelberg, pp. 216-222.

SCHWARZ, H. (1975) : Stichprobenverfahren, Oldenbourg, München und Wien.

STADLER, R. (1989) : Stand, Entwicklungsperspektiven und Grenzen der Datenbeschaffung und Erhebungsmethoden in der Amtlichen Agrarstatistik, Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften, Münster-Hiltrup, Vol. 25, pp. 153-156.

STADLER, R. (1991) : Satellitenfernerkundung und amtliche Statistik, Baden- Württemberg in Wort und Zahl, Statistisches Landesamt Baden-Württemberg, Band 39, Heft 3, pp. 127-141.

STADLER, R. (1992) : Fortschritte bei der Flächen- unde Ertragsermittlung via Satelliten-Fernerkundung, Baden-Württemberg in Wort und Zahl, Statistisches Landesamt Baden-Württemberg, Band 40, Heft 9, pp. 431-441.

STENGER, H. (1977) : Antwortvariabilität als Fehlerquelle, Allgemeines Statistisches Archiv, Band 61, Heft 4 pp. 391-403.

STRECKER, H., WIEGERT, R., PEETERS, J., KAFKA, K. (1983) : Messung der Antwortvariabilität auf Grund von Erhebungsmodellen mit Wiederholungszählungen, Schriftenreihe : Angewandte Statistik und Ökonometrie, Heft 25, Vandenhoeck und Ruprecht, Göttingen.

STRECKER, H., WIEGERT, R. (1984) : Der Mean Square Error und die Antwortvarianz bei Erhebungen - Darstellung und Schätzung, Allgemeines Statistisches Archiv, Band 68, pp. 199-222.

STRECKER, H., PEETERS, J., KAFKA, K., WIEGERT, R. (1985) : L'enquête de contrôle, un instrument permettant de déceler les erreurs dans les déclarations des effectifs et de déterminer la variance des réponses. Recensement agricole et horticole au 15 mai 1979 en Belgique, in : Etudes statistiques, n° 75, Institut national de statistique, Bruxelles, 56 p.

STRECKER, H., WIEGERT, R. (1986) : Die Antwortvariabilität bei statistischen Erhebungen - Wiederholungszählungen und Schätzung der Antwortvarianz, Österreichische Zeitschrift für Statistik und Informatik, 16. Jhg., pp. 99-130.

STRECKER, H., WIEGERT, R. (1989) : Wirtschaftsstatistische Daten und ökonomische Realität, Jahrbücher für Nationalökonomie und Statistik, Band 206, pp. 487-509.

STRECKER, H., (1996) : Zur Fehlermessung bei nach Grössenklassen aufgegliederten Erhebungsergebnissen, Der Inkonsistenz-Index, Jahrbücher für Nationalökonomie und Statistik, Band 215, Heft 2, pp. 216-227.

STRECKER, H., (1997) : Der Inkonsistenz-Index - Eine Masszahl zur Fehlermessung bei nach Grössenklassen aufgegliederten Erhebungsergebnissen, Jahrbücher für Nationalökonomie und Statistik, Band 216, Heft 6, pp. 625-657; Band 217, Heft 2, p. 184.

STRECKER, H., (1999) : Inconsistency, Strecker's Index of Encyclopedia of Statistical Sciences, Update Vol. III (Eds. S. Kotz, C.B. Read, D.L. Banks), John Wiley & Sons, New York and London, pp. 359-361.

STRECKER, H., (1980) : Model for the Decomposition of Errors in Statistical Data into Components and the Ascertainment of Response Errors by Means of Accuracy Checks, Jahrbücher für Nationalökonomie und Statistik, Band 195, Heft 5, pp. 385-420.

SUKHATME, P.V., SUKHATME, B.V., SUKHATME, S., ASOK, C. (1984) : Sampling Theory of Surveys with Applications, IOWA-Press, pp. 487-506.

U.S. Bureau of the Census (1960) : Evaluation and Research Program of the U.S. Census of Population and Housing 1960 - The employer Record check, Series ER 60, No. 6, 14 p.

WARNER, S.L. (1965) : Randomized response : A survey technique for eliminating evasive answer bias, Journal of the American Statistical Association, Vol. 60, pp. 63-69.

WEICHSELBERGER, K. (1970) : Genauigkeitsansprüche, Fehler und Kontrollen bei Volkszählungen, Allgemeines Statistisches Archiv, Band 54, Heft 1, pp. 1-28.

WIEGERT, R. (1982) : Überlegungen zum Adäquations- und Fehlerbegriff in Statistik und Ökonometrie, in : Piesch, W., Förster, W. (ed.), Angewandte Statistik und Wirtschaftsforschung heute, Reihe Angewandte Statistik und Ökonometrie, Nr. 21, Vandenhoeck und Ruprecht, Göttingen, pp. 254-266.

ZARKOVICH, S.S. (1966) : Quality of Statistical Data, FAO, Roma.

ANNEXES

MINISTERE DES AFFAIRES ECONOMIQUES

1000 BRUXELLES, le

Rue de Louvain, 44

Tél. 02 / 513.96.50

PET03F-018

Institut National de Statistique

STATISTIQUES AGRICOLES

Votre lettre du	Vos références	Nos références	Annexes
		7.LA/65/85-2/656	

Objet : Enquête-pilote au 4 mai dans le cadre du recensement agricole et horticole au 15 mai 1985.

Mademoiselle, Madame, Monsieur,

Vous êtes agriculteur et je vous saurais gré de bien vouloir collaborer à l'enquête-pilote susmentionnée.

Cette enquête a pour but de suivre l'évolution qui se produit à court terme pour certaines cultures e.a. le froment d'hiver, le froment de printemps, le seigle d'hiver, les betteraves sucrières, etc.

L'analyse de ce changement est d'une très grande valeur notamment en ce qui concerne la prévision de l'offre sur le marché belge.

Dans ce domaine, cette enquête poursuit un test à l'aide d'un questionnaire simplifié.

Un certain nombre d'exploitations dont la vôtre a été sélectionnée selon la méthode de l'échantillonnage aléatoire.

Puis-je donc vous demander de remplir soigneusement le questionnaire au verso de cette lettre.

La situation des cultures à déclarer (céréales, cultures industrielles et récapitulation générale) est celle existante au 4 mai 1985.

Je puis vous garantir formellement que les informations recueillies seront dépouillées de façon strictement confidentielles au niveau du Royaume.

Pour le renvoi du questionnaire, de préférence par retour du courrier, veuillez utiliser l'enveloppe pré-affranchie ci-jointe.

Si vous constatez certaines erreurs dans le libellé de votre adresse, veuillez les corriger.

Je vous remercie d'avance pour votre précieuse collaboration.

Pour le Directeur général :
Le Statisticien principal-Chef de service ff.,

J. PEETERS.

Année de naissance ..... 281 |__|__|__|__|

CULTURES PRATIQUEES EN CULTURE PRINCIPALE

	Code	ha	a
Céréales pour le grain (y compris semences)			
Froment d'hiver	014		
Froment de printemps	015		
Seigle d'hiver	016		
Epeautre	017		
Orge d'hiver(escourgeon)	018		
Orge de printemps	019		
Avoine	020		
Maïs cultivé pour la graine	021		
Triticale (1)	022		
Autres céréales et mélanges de céréales	023		
TOTAL (codes 014 à 023)	025		
Plantes industrielles			
Betteraves sucrières (excepté les semences)	026		
Lin	027		
Chicorée à café (excepté les semences)	028		
Colza d'hiver	029		
Colza d'été	030		
Autres plantes oléagineuses	031		
Tabac	032		
Houblon	033		
Plantes médicinales,aromatiques et condimentaires	034		
TOTAL (codes 026 à 034)	038		

RECAPITULATION GENERALE

	Code	ha	a
Prés et prairies (code 007)	240		
Légumes à cosse récoltés secs (code 013)	241		
Céréales (code 025)	242		
Plantes industrielles (code 036)	243		
Semences agricoles et plants pour la vente (code 046)	244		
Plantes racines et tuberculifères (code 051)	245		
Fourrages verts (code 060)	246		
Pommes de terre (code 065)	247		
Cultures agricoles non dénommées (code 066)	248		
Fleurs, bulbes à fleurs, fleurs à couper et plantes ornementales (code 080)	249		
Cultures fruitières (code 238)	250		
Arbres et arbustes de pépinières(code 094)	251		
Cultures de légumes : extensives (code 178)	252		
intensives (code 180)	253		
Installations sous verre (code 181)	254		
Semences horticoles, plants de légumes et plants de fleurs (code 222)	255		
Cultures horticoles pour la consommation du ménage du déclarant (code 227)	256		
Oseraies (code 228)	257		
Terre en repos (code 229)	258		
SUPERFICIE AGRICOLE UTILISEE (code 240 à 258)	260		

Certifié exact et sincère, le.....198

Le déclarant

MINISTERE DE L'AGRICULTURE
MINISTERE DES AFFAIRES ECONOMIQUES
INSTITUT NATIONAL DE STATISTIQUE

Bureau pour la province de Brabant
Rue de Louvain 44 - 1000 Bruxelles

MODELE 1

RECENSEMENT AGRICOLE ET HORTICOLE AU 15 MAI 1985

TRES IMPORTANT : Avant de remplir le questionnaire, veuillez le lire attentivement ainsi que les instructions complémentaires, vous éviterez de cette façon des inscriptions erronées, vous faciliterez votre travail ainsi que le nôtre.
En vertu de la loi, l'utilisation, dans un but fiscal, des renseignements individuels fournis à l'occasion de ce recensement est strictement défendue.
Les numéros entre parenthèses, par exemple (1), renvoient aux instructions.
A certaines questions, il y a lieu de répondre en traçant une croix dans la case appropriée, par exemple ☒.

EMPLACEMENT DE L'EXPLOITATION

Commune de situation du siège de l'exploitation .....
(emplacement des principaux bâtiments de l'exploitation)

Dans le cas où il n'existe pas de siège d'exploitation, la déclaration est recueillie à la commune de résidence du déclarant

DOIVENT REPONDRE AU PRESENT QUESTIONNAIRE :
(marquer d'une ☒ la case correspondant à votre situation)

1. toutes les personnes dont la profession principale (profession qui occupe la plus grande partie du temps, dans les cas douteux celle qui rapporte le plus) est agriculteur ou éleveur. ..... ☐ 1
2. toutes les personnes dont la profession principale (profession qui occupe la plus grande partie du temps, dans les cas douteux celle qui rapporte le plus) est horticulteur. ..... ☐ 2
3. tous les exploitants autres que ceux visés sous 1 et 2 produisant en vue de vendre des produits végétaux ou animaux non compris les personnes exécutant des travaux à l'entreprise. ..... ☐ 3
4. tous les établissements pénitentiaires, pensionnats, maisons de repos, hospices et autres établissements similaires, les communautés religieuses ou autres, les établissements d'expérimentation ou de recherche, les services de plantations d'organismes publics produisant pour leurs propres besoins et les exploitations dépendant d'un établissement d'instruction, pour autant qu'ils exploitent au moins un are ou détiennent des animaux mentionnés dans le questionnaire, qu'ils produisent ou non pour la vente ..... ☐ 4
5. toutes les personnes ou organismes exécutant des travaux agricoles ou horticoles à titre d'activité principale ou accessoire, pour le compte d'agriculteurs ou d'horticulteurs ou mettant des machines agricoles ou horticoles ou des installations à leur disposition et qui produisent également en vue de vendre, des produits végétaux ou animaux ..... ☐ 5
6. toutes les personnes ou organismes exécutant des travaux agricoles ou horticoles à titre d'activité principale ou accessoire, pour le compte d'agriculteurs ou d'horticulteurs ou mettant des machines agricoles ou horticoles ou des installations à leur disposition, pour autant qu'ils ne produisent pas en vue de vendre, des produits végétaux ou animaux. ..... ☐ 6
7. toutes les coopératives d'utilisation en commun de machines agricoles ou horticoles. ..... ☐ 7

n.b. 1) Contrat: Dans les cas de production sous contrat, doivent répondre :
pour les produits végétaux : les agriculteurs et les horticulteurs - professionnels ou non - qu'ils récoltent ou non le produit; pour les produits animaux : les détenteurs des animaux, c.à.d. les personnes chez qui les animaux se trouvent.
2) Dans les cas de cultures associées c.à.d effectuées simultanément sur une même superficie et fournissant normalement chacune une production bien distincte, il faut répartir cette superficie de telle façon que la superficie réelle consacrée à chacune des cultures puisse être déclarée.

IDENTIFICATION DU CHEF D'EXPLOITATION (pour les nouveaux déclarants, prière d'écrire en caractères d'imprimerie)

(personne qui a la responsabilité de la gestion journalière de l'exploitation. Si plusieurs personnes assurent en commun la gestion de l'exploitation, mentionner uniquement le nom du principal responsable ou de la personne la plus âgée).

PERSONNALITE JURIDIQUE DE L'EXPLOITATION

	Code	
Le responsable juridique est-il : -une personne morale (société, communauté religieuse, centre publique d'aide sociale)?		☐ 1
-une personne physique?	280	☐ 2
Si le responsable juridique est une personne physique, est-il en même temps chef d'exploitation?	279	☐ 3

Année de naissance	281	____
Etes-vous devenu chef d'exploitation entre le 15 mai 1984 et le 15 mai 1985 ?	282	1 ☐ oui 2 ☐ non
Dans la négative, depuis quand êtes-vous chef d'exploitation ?	283	____

Profession principale du chef d'exploitation(occupation qui absorbe la plus grande partie du temps; dans les cas douteux, celle qui procure le revenu le plus important): .....

Réservé à l'I.N.S. Code 996 .....

CULTURES PRATIQUEES EN CULTURE PRINCIPALE

	Code	ha	a
Prés et prairies			
Prairies temporaires (occupant des terres pendant une ou seulement quelques années) à faucher:			
de ray-grass italien et de Westerwold (en semis pur)	001		
autres (dont trèfles en mélange avec éventuellement céréales en vert comme plante abri)	002		
Prairies permanentes (à l'exclusion de tous prés-vergers)			
- exclusivement pour la pâture	004		
- exclusivement pour la fauche	005		
- mixte (pour la pâture et la fauche)	006		
Sous-total (codes 005 + 006)	003		
TOTAL (codes 001 + 002 + 003 + 004)	007		
Légumes à cosse récoltés secs (y compris semences)			
Haricots secs	008		
Pois secs	009		
Autres (y compris mélanges de céréales et de légumes, pour la graine)	010		
TOTAL (codes 008 à 010)	013		
Céréales pour le grain (y compris semences)			
Froment d'hiver	014		
Froment de printemps	015		
Seigle d'hiver	016		
Epeautre	017		
Orge d'hiver(escourgeon)	018		
Orge de printemps	019		
Avoine	020		
Maïs cultivé pour la graine	021		
Triticale (1)	022		
Autres céréales et mélanges de céréales	023		
TOTAL (codes 014 à 023)	025		
Plantes industrielles			
Betteraves sucrières (excepté les semences)	026		
Lin	027		
Chicorée à café (excepté les semences)	028		
Colza d'hiver	029		
Colza d'été	030		
Autres plantes oléagineuses	031		
Tabac	032		
Houblon	033		
Plantes médicinales,aromatiques et condimentaires	034		
TOTAL (codes 026 à 034)	038		
Plants et semences agricoles pour la vente (non compris les semences de céréales)			
Plants de pommes de terre	039		
Semences de graminées	040		
Semences de betteraves sucrières	041		
Autres semences	042		
TOTAL (codes 039 à 042)	046		

	Code	ha	a
Cultures fourragères:			
a) Plantes racines et tuberculifères (en culture principale) :			
Betteraves fourragères et mi-sucrières (excepté les semences)	047		
Toutes autres plantes racines et tuberculifères	048		
TOTAL (codes 047 + 048)	051		
b) Fourrages verts (excepté pour la production de semences):			
Trèfles	052		
Luzerne	053		
Maïs laiteux ou pâteux	054		
Autres fourrages verts et mélanges (excepté les prés et prairies)	055		
c) Maïs dont le grain ou l'épi sont conservés à l'état humide	057		
TOTAL (codes 052 à 055 + 057)	060		
Pommes de terre (excepté les cultures de plants)			
Pommes de terre hâtives (récoltées avant le 1er août)	061		
Pommes de terre mi-hâtives	062		
Pommes de terre mi-tardives et tardives	063		
TOTAL (codes 061 à 063)	065		
Cultures agricoles non spécifiées ci-dessus	066		
Cultures de plein air de fleurs, de bulbes à fleurs, de fleurs à couper et de plantes ornementales pour la vente :			
Superficie plantée ou réservée à la plantation de:			
Fleurs à couper	070		
Chrysanthèmes en pot	071		
Plantes vivaces (rustiques) et plantes en massif	072		
Bégonias	073		
Azalées	074		
Tulipes (pour le bulbe)	075		
Autres bulbes et tubercules à fleurs	076		
Autres plantes ornementales pour la fleur et/ou le feuillage	077		
TOTAL (codes 070 à 077)	080		
Cultures de plein air de semences horticoles, de plants de légumes et de plants de fleurs pour la vente (y compris les cultures qui au 15 mai sont sous cloches, petits tunnels et plastique perforé).			
Plants de fraisiers	216		
Semences et plants de fleurs et de plantes d'ornement	217		
Autres semences et plants horticoles	218		
TOTAL (codes 216 à 218)	222		
Cultures de plein air d'arbres et d'arbustes de pépinières destinés à la vente (indiquer la superficie totale et détailler dans le questionnaire Mod. III)	094		

CULTURES DE CHAMPIGNONS (superficie des couches)	Code	m2
Dans des bâtiments spécialement aménagés	231	
En grottes, carrières, caves	232	

CULTURES PRATIQUEES EN CULTURE PRINCIPALE

Cultures de légumes en plein air pour la vente * (étendues occupées par les cultures, ou, si la terre est libre le 15 mai, étendues éventuellement destinées à la plantation et/ou au repiquage de ces cultures).	Cultures extensives (en assolement avec des cultures agricoles)			Cultures intensives (en rotation avec d'autres cultures de légumes)			
	Code	ha	a	Code	ha	a	ca (m2)
Pois verts pour l'industrie de la conserve	100			101			
Pois verts pour consommation à l'état frais	102			103			
Haricots nains récoltés pour l'industrie de la conserve (1)	104			105			
Haricots nains récoltés pour consommation à l'état frais (1)	106			107			
Haricots à rames (1)	108			109			
Autres haricots verts (1)	110			111			
Carottes potagères pour l'industrie de la conserve (1)	112			113			
Carottes potagères pour consommation à l'état frais (1)	114			115			
Céleris blancs pour l'industrie de la conserve (1)	116			117			
Céleris blancs pour consommation à l'état frais (1)	118			119			
Epinards pour l'industrie de la conserve (1)	120			121			
Epinards pour consommation à l'état frais (1)	122			123			
Cerfeuil pour l'industrie de la conserve (1)	124			125			
Cerfeuil pour consommation à l'état frais (1)	126			127			
Oignons : Petits oignons blancs (pour conserve)	128			129			
Autres (semences - plants et bulbes à planter)	130			131			
Chicorées Witloof pour la production de chicons :							
avec forçage à l'exploitation	132			133			
sans forçage à l'exploitation	134			135			
Tomates	136			137			
Choux-fleurs (1)	138			139			
Brocoli	140			141			
Asperges	142			143			
Laitues pommées (1)	144			145			
Choux de Bruxelles (1)	146			147			
Choux rouges (1)	148			149			
Choux blancs (1)	150			151			
Choux de Savoie (1)	152			153			
Poireaux (1)	154			155			
Echalotes	156			157			
Céleris raves (1)	158			159			
Céleris verts (1)	160			161			
Scorsonères	162			163			
Fenouil	164			165			
Chicorées frisées et scaroles (1)	166			167			
Persil	168			169			
Radis	170			171			
Rhubarbe	172			173			
Cornichons	174			175			
Autres légumes (non compris les fraises)	176			177			
TOTAL (codes 100 à 176 et 101 à 177)	178			179			
				Superficie (code 179) (arrondie à l'are)	180		

(1) Récoltés normalement avant le 1er septembre
* Y compris les cultures qui au 15 mai sont sous cloches, petits tunnels et plastique perforé.

FORCAGE DE WITLOOF

Ne doivent répondre à cette rubrique que les exploitants produisant des racines et pratiquant le forçage sur l'exploitation.

	Code	m2
Superficie des couches	215	

Après les cultures énumérées ci-dessus, avez-vous l'intention de procéder à des cultures de légumes de plein air que l'on trouve normalement en place le 1er septembre ou les jours qui suivent?

	Code	
oui	230	1 ☐
non		2 ☐

INSTALLATION SOUS VERRE OU MATIERE PLASTIQUE :

serres, warenhuis, grands tunnels et châssis portables (à l'exclusion des petits tunnels et cloches) :

	Code	ha	a
Superficie au sol (pour installations sur rail, toute la superficie pouvant être couverte)	181		
-dont - superficie chauffée	182		
- superficie non chauffée	183		

DETAIL DES SUPERFICIES DES CULTURES AU 15 MAI, sous serres, warenhuis, grands tunnels et châssis portables y compris les superficies des tablettes fixes ou suspendues mais à l'exclusion des petits tunnels et cloches:	Code	m2
Semences de légumes	184	
Semences de fleurs et de plantes d'ornement	185	
Plants de légumes	186	
Plants pour parterres, jardinières, balcons et terrasses	187	
CULTURES DE LEGUMES:		
Haricots	188	
Céleris blancs	189	
Céleris verts	190	
Cornichons	191	
Tomates: sous verre chaud	192	
sous verre froid	193	
Concombres: sous verre chaud	194	
sous verre froid	195	
Autres légumes, y compris melons: sous verre chaud	196	
sous verre froid	197	
CULTURES DE FLEURS:		
Azalées	198	
Bulbes et tubercules	199	
Plantes en pot (plantes d'appartement)	200	
Fleurs à couper: Roses	201	
Oeillets	202	
Chrysanthèmes	203	
Autres	204	
Autres espèces de fleurs et plants	205	
Pépinières sous verre et/ou matière plastique	206	
CULTURES FRUITIERES:		
Raisins: Royal	207	
Muscat	208	
Autres	209	
Fraises: forçage intensif	210	
forçage léger	211	
cultures froides	212	
Autres sortes de fruits	213	
TOTAL (codes 184 à 213)	214	

Cultures horticoles de plein air et sous verre pratiquées exclusivement pour la consommation du ménage du déclarant (cat. 4 et établissements spéciaux: déclaration à faire sous les codes 249 à 255 de la récapitulation générale p. 5).	Code	m2
Jardins potagers	223	
Cultures fruitières (y compris les prés-vergers) (non compris sous le code 238)	224	
TOTAL (codes 223 + 224)	226	

		ha	a
A reporter ci-contre, chiffres arrondis à l'are (code 226)	227		
Oseraies	228		
Terres labourées, temporairement en repos et ne devant pas porter de culture avant l'automne prochain	229		

CULTURES FRUITIERES DE PLEIN AIR (Pour la vente)

Vergers à hautes tiges

Y compris les prés-vergers	Code	ha	a	Code	Nombre d'arbres
Pommes: Belle de Boskoop	081			765	
Autres	082			766	
Poires: Légipont	083			767	
Autres	084			768	
Cerises	085			769	
Prunes	086			770	
Autres espèces et variétés	087			771	
Vergers à basses tiges					
Pommes: Golden Delicious	088			772	
Boskoop	089			773	
Cox's Orange Pippin	090			774	
Jonagold	091			775	
Autres	092			776	
Poires: Conférence	093			777	
Doyenné du Comice	095			778	
Durondeau	096			779	
Autres	097			780	
Cerises: Douces	098			781	
Aigres	099			782	
Prunes	219			783	
Pêches	220			784	
Autres espèces et variétés	221			785	
Autres cultures fruitières de plein air : Fraises: -en pleine terre	233				
-sous petits tunnels, cloches ou plastique perforé	234				
Framboises	235				
Groseilles rouges	236				
Autres baies (cassis, mûres, groseilles vertes, etc...)	237				
TOTAL (codes 081 à 237)	238				

RECAPITULATION GENERALE	Code	ha	a
Prés et prairies (code 007)	240		
Légumes à cosse récoltés secs (code 013)	241		
Céréales (code 025)	242		
Plantes industrielles (code 038)	243		
Semences agricoles et plants pour la vente (code 046)	244		
Plantes racines et tuberculifères (code 051)	245		
Fourrages verts (code 060)	246		
Pommes de terre (code 065)	247		
Cultures agricoles non dénommées (code 066)	248		
Fleurs, bulbes à fleurs, fleurs à couper et plantes ornementales (code 080)	249		
Cultures fruitières (code 238)	250		
Arbres et arbustes de pépinières(code 094)	251		
Cultures de légumes : extensives (code 178)	252		
intensives (code 180)	253		
Installations sous verre (code 181)	254		
Semences horticoles, plants de légumes et plants de fleurs (code 222)	255		
Cultures horticoles pour la consommation du ménage du déclarant (code 227)	256		
Oseraies (code 228)	257		
Terre en repos (code 229)	258		
SUPERFICIE AGRICOLE UTILISEE (code 240 à 258)	260		

AUTRES PARTIES DE L'EXPLOITATION	Code	ha	a
Superficie agricole non utilisée (2)	261		
Superficie boisée	262		
Autres superficies (3)	263		
Superficie de sapins de Noël (non compris au code 262)	264		
Superficie totale de l'exploitation (codes 260 à 264)	270		

MODE DE FAIRE-VALOIR			
Superficie agricole utilisée :			
en faire-valoir direct (4)	271		
en fermage (5)	272		
en métayage et en autres modes de faire-valoir (6)	273		
TOTAL (codes 271, 272 et 273; doit correspondre au code 260)	274		

IRRIGATION (par aspersion et/ou par gravité)			
Superficie normalement irriguée	291		
Superficie irrigable avec les moyens techniques disponibles sur l'exploitation	292		
Si vous n'irriguez pas, envisagez-vous d'irriguer à l'avenir ?			
Oui	293	☐	1
Non		☐	2

PRODUCTION	COMMERCIALISATION	Code	Sous contrat	Pour compte propre	Mixte
PORCINS	- élevage	396	☐ 1	☐ 2	☐ 3
Votre porcherie est-elle du type	- engraissement	397	☐ 1	☐ 2	☐ 3
	- mixte (élevage + engraissement)	398	☐ 1	☐ 2	☐ 3

PLANTATIONS FRUITIERES

Avez-vous apporté au cours de la campagne 1984-1985 des modifications à vos plantations fruitières à basses tiges soit par arrachage, soit par regreffage ?

	Code	
oui	787	1 ☐
non		2 ☐

Si oui, prière de remplir le questionnaire spécial Modèle II.

Avez-vous procédé à de nouvelles plantations, fruitières à basses tiges au cours de la campagne 1984-1985 ?

	Code	
oui	788	1 ☐
non		2 ☐

Si oui, prière de remplir le questionnaire spécial Modèle II.

HABITATION	Code	Année
Habitation complète mise en service en	949	

Réservé à l'I.N.S. Code 997 .....

ANIMAUX

	Code	Nombre
Bovins		
A. Bovins de moins de 1 an: destinés à être abattus comme veaux	300	
autres: mâles	301	
femelles	302	
B. Bovins de 1 à moins de 2 ans:		
mâles	304	
femelles:		
animaux de boucherie	305	
primipares à l'engrais (moins de 2 ans)	306	
autres (destinées à remplacer les vaches visées sous les codes 313 et 314)	307	
C. Bovins de 2 ans et plus:		
mâles	308	
génisses:		
animaux de boucherie	309	
primipares à l'engrais (2 ans et plus)	310	
autres (destinées à remplacer les vaches visées sous les codes 313 et 314)	311	
vaches, y compris vaches de réforme, mais non compris primipares à l'engrais dont:		
pour la traite	313	
pour la spéculation du veau au pis (vaches non traites)	314	
TOTAL (codes 300 à 314)	320	
Répartition des vaches d'après la race et le croisement (codes 313 et 314)		
A. Races reconnues (7)		
1. Blanc-bleu de Belgique	321	
2. Pie-rouge de Belgique	322	
3. Pie-noire de Belgique	323	
4. Rouge de Flandre occidentale	324	
5. Blanc-rouge de Flandre orientale	325	
B. Races admises		
1. Charolaise	326	
2. Limousine	327	
3. Holstein	328	
4. Blonde d'Aquitaine	329	
5. Jersey	330	
C. Autres races	331	
D. Croisements		
1. pour la viande	332	
2. pour le lait	333	
TOTAL (codes 321 à 333, doit correspondre aux codes 313 + 314)	335	
Moutons		
De moins de 1 an :		
Mâles	371	
Femelles : saillies	372	
non saillies	373	
De 1 an et plus :		
Mâles	374	
Femelles : saillies	378	
non saillies	379	
TOTAL (codes 371 à 379)	375	
Chèvres	376	

	Code	Nombre
Porcs		
A. Porcelets d'un poids vif de moins de 20 kg	340	
B. Porcs d'un poids vif de 20 kg à moins de 50 kg	341	
C. Porcs à l'engrais (y compris les verrats de réforme et les truies de réforme) d'un poids vif:		
a) de 50 kg à moins de 80 kg	342	
b) de 80 kg à moins de 110 kg	343	
c) de 110 kg et plus	344	
D. Porcs reproducteurs d'un poids vif de 50 kg et plus:		
a) verrats	345	
b) truies saillies:		
1. truies saillies pour la première fois	346	
2. autres truies saillies	347	
c) truies d'élevage non saillies:		
1. jeunes truies non encore saillies	348	
2. autres truies	349	
TOTAL (codes 340 à 349)	355	
Répartition des truies d'après la race ou le croisement (8)		
1. Landrace	390	
2. Piétrain	391	
3. Large White	392	
4. Hybrides	393	
5. Autres races	394	
TOTAL (codes 390 à 394; doit correspondre aux codes 346 à 349)	395	
Volaille		
Poules et poulettes non en âge de ponte	380	
Poules pondeuses en âge de ponte, y compris poules de réforme	381	
Coqs pour la reproduction	382	
Poulets de chair (non compris les poussins d'un jour)	383	
Canards	384	
Oies	385	
Dindons	386	
Pintades	387	
Apiculture		
Ruches en paille	358	
Caisses à cadres mobiles	359	
Chevaux agricoles (chevaux utilisés principalement ou accessoirement aux travaux de l'exploitation):		
de moins de 3 ans	360	
de 3 ans et plus	361	
TOTAL (codes 360 + 361)	362	
Autres chevaux (uniquement chevaux de voiture, de selle, de course): chevaux	363	
poneys	364	
TOTAL (codes 363 + 364)	365	
Anes, mulets et bardots	370	
Lapins	377	

Réservé à l'I.N.S. Code 998

MATERIEL AGRICOLE

Les machines en copropriété doivent être déclarées par le copropriétaire qui les détient à la date du 15 mai. Les machines qui, à la date du 15 mai, ne se trouvent pas à l'exploitation pour quelque motif que ce soit (par exemple: réparation, prêt, location) doivent être déclarées à l'exploitation dont elles relèvent ou, en cas de copropriété qui les détient normalement.

Les machines appartenant à des entrepreneurs de travaux ou à des coopératives qui, au 15 mai, sont utilisées dans une exploitation donnée, doivent être déclarées par les entrepreneurs de travaux ou les coopératives.

	Code	Nombre
Tracteurs : - 2 roues motrices	430	
- 4 roues motrices	431	
Motoculteurs (tous types) (9)	432	
Moissonneuses-batteuses (tous types)	433	
Récolteuses-hacheuses (tous types)	434	
Distributeurs d'engrais solide		
a) à projection		
tractés	435	
portés	436	
b) sans projection		
trémie de moins de 5 tonnes	437	
trémie de plus de 5 tonnes	438	
c) pneumatiques	439	

	Code	Nombre
Installations de traite mécanique		
a) à l'exploitation	440	
- sans lactoduc (canal à lait)	441	
- avec lactoduc (canal à lait)	442	
b) en prairie	443	
Tanks refroidisseurs de lait	444	
Installations de séchage de fourrages	445	
Moulins, concasseurs, applatisseurs de grains	446	
Mélangeurs d'aliments pour bétail	447	
Systèmes électroniques de distribution d'aliments pour le bétail	448	
Engins automoteurs de manutention (électriques ou autres) (10)	449	
Pompes à lisier	450	
Silos à lisier	451	
Installations d'épuration du lisier	452	

SILOS

(pour les silos taupinières, indiquer l'ensilage réalisé pendant la saison culturale 1984 - 1985)

Indication	Type de silo	TAUPINIERES (capacité = volume après tassement) (12)	SILOS COULOIRS avec parois en (11)						SILOS-TOURS			
			béton coulé sur place		maçonnerie ou béton préfabriqué		bois		en métal		autres	
ENSILAGE DE MAIS	Nombre de silos	510	511		512		513		514		515	
	Capacité (m3)	516	517		518		519		520		521	
ENSILAGE D'HERBE	Nombre de silos	522	523		524		525		526		527	
	Capacité (m3)	528	529		530		531		532		533	
AUTRES ENSILAGES	Nombre de silos	534	535		536		537		538		539	
	Capacité (m3)	540	541		542		543		544		545	

DESILEUSES POUR SILOS-TOURS

Par le haut	mecanique		☐	1		☐	1		☐	1		☐	1		☐	1
	pneumatique	546	☐	2	547	☐	2	548	☐	2	549	☐	2	550	☐	2
Par le bas			☐	3		☐	3		☐	3		☐	3		☐	3

SILOS POUR ALIMENTS SECS (concentrés ou farineux)

Nombre de silos		Capacité totale		Servant au stockage d'aliments pour			
Code	Nombre	Code	m3	Code	bovins	porcs	volailles
551		552		553	☐ 1	☐ 2	☐ 3

DESILEUSES POUR SILOS HORIZONTAUX (13)

Mise en service en		Traction				Type		
Code	Année	Code	portée	semi-portée	tractée	Code	en bloc	en vrac
554	19 . .	555	☐ 1	☐ 2	☐ 3	556	☐ 1	☐ 2

TONNES A LISIER

	Mise en service		Capacité		Type			
	Code	Année	Code	m3	Code	Sans pompe	Avec pompe à vide	Avec pompe à vis
1ère tonne	557	19 . .	558	.....	559	☐ 1	☐ 2	☐ 3
2ème tonne	560	19 . .	561	.....	562	☐ 1	☐ 2	☐ 3

TABLE AUTOMATIQUE DE DISTRIBUTION D'ALIMENTS

Longeur totale		Type				
Code	Mètres	Code	va et vient	à tapis	à raclettes	à soufflerie
563	.....	564	☐ 1	☐ 2	☐ 3	☐ 4

EVACUATEUR DE DEJECTIONS

Type	Mise en service		Longueur totale	
	Code	Année	Code	Mètres
A raclette va et vient	565	.....	566	.....
A raclette continue	567	.....	568	.....
A raclette pliante (en V)	569	.....	570	.....

ETABLES POUR BOVINS (14)

NOMBRE DE BATIMENTS	
Code	
600	.....

RECENSER UNIQUEMENT LES ETABLES AYANT UNE CAPACITE (exprimée en animaux logeables) DE 15 VACHES ET PLUS (laitières ou allaitantes) OU DE 40 BOVINS A L'ENGRAIS ET PLUS OU DE 50 VEAUX A L'ENGRAIS ET PLUS.

NE DOIVENT PAS ETRE RECENCES : Les infirmeries, les maternités, les étables de quarantaine ou désaffectées.
Les locaux (souvent anciennes étables) qui sont OCCASIONNELLEMENT affectés au logement MOMENTANE de bovins.
Les locaux pour le jeune bétail d'élevage et/ou de remplacement (futurs reproducteurs/trices).

NOTICE

1) Si un exploitant possède deux étables de même type, mises en service à des époques différentes, indiquer seulement l'année de mise en service de la dernière étable construite.
Dans ce cas, toutefois on recensera les caractéristiques techniques (locaux, animaux logeables, etc.) des deux étables.

2) Si plusieurs sortes d'animaux sont logées dans un même local (p. ex. vaches laitières et vaches allaitantes ou vaches laitières et bovins à l'engrais) on considérera que chaque sorte d'animaux est logée dans un local séparé, à recenser indépendamment.

3) Année de mise en service d'une étable (ou d'une portion d'étable) signifie : année de 1ère mise en service (ou de transformation radicale) dans la fonction actuelle.
EXEMPLE :
Etable à stabulation libre semi-paillée, pour bovins à l'engrais, construite en 1972 et dont une moitié aurait été convertie en 1979, en stabulation libre à logettes pour vaches allaitantes.
Code 683 = 1 Code 682 = 1972
Code 611 = 1 Code 610 = 1979

4) a) Indiquer seulement le nombre de vaches laitières logeables.
b) Indiquer seulement le nombre de vaches allaitantes logeables.

5) La notice aux agents recenseurs comporte de plus amples explications relatives aux salles de traite et aux étables.

SALLES DE TRAITE (pour les étables laitières à stabulation libre : 1.1; 1.2; 1.3)							
Année de mise en service	Type d'installation (14)	en arête de poisson	en tandem	en tunnel	en stalles parallèles	en carrousel
Code/601 19 . .	Nombre de stalles disponibles	602	603	604	605	606

Etables (14) / Spécifications de locaux	Capacité de logement et évacuation des déjections	Année de mise en service		Locaux		Animaux logeables		Evacuation des déjections: caillebotis		tracteur		raclette		manuelle	
		Code	Année	Code	Nombre	Code	Nombre	Code	Nombre de locaux	Code	Nombre de locaux	Code	Nombre de locaux	Code	Nombre de locaux
1. ETABLES POUR BETAIL LAITIER : (a)															
1.1 Stabulation libre à logettes															
Alimentation :															
– dans l'étable															
– à la crèche		610		611		612		613		614		615			
– à la logette		616		617		618		619		620		621			
– à l'extérieur de l'étable		622		623		624		625		626		627			
1.2. Stabulation libre paillée		628		629		630									
1.3. Stabulation libre semi-paillée		631		632		633		634		635		636			
1.4. Stabulation entravée															
a) à grilles avec :															
– aération naturelle		640		641		642									
– ventilation forcée		643		644		645									
b) à court bâti :															
– aération naturelle		646		647		648				649		650		651	
– ventilation forcée		652		653		654				655		656		657	
dont : sans couloir d'alimentation		658		659		660									
c) à long bâti (ou semi-long)		661		662		663									
2. ETABLES POUR BOVINS A L'ENGRAIS (autres que veaux abattus jeunes)															
2.1. Stabulation sur caillebotis total		670		671		672									
2.2. Stabulation libre à logettes		673		674		675		676		677		678			
2.3. Stabulation libre paillée		679		680		681									
2.4. Stabulation libre semi-paillée		682		683		684		685		686		687			
2.5. Stabulation entravée avec :															
– aération naturelle		688		689		690		691		692		693		694	
– ventilation forcée		695		696		697		698		699		700		701	
3. ETABLES POUR VACHES ALLAITANTES (b)															
3.1. Stabulation libre à logettes		710		711		712		713		714		715			
3.2. Stabulation libre paillée		716		717		718									
3.3. Stabulation libre semi-paillée		719		720		721		722		723		724			
3.4. Stabulation entravée :															
a) avec veaux attachés :															
– aération naturelle		725		726		727		728		729		730		731	
– ventilation forcée		732		733		734		735		736		737		738	
b) avec veaux non attachés :															
– aération naturelle		739		740		741		742		743		744			
– ventilation forcée		745		746		747		748		749		750			
4. ETABLES POUR VEAUX ABATTUS JEUNES		760		761		762									

(a), (b) Voir notice explicative ci-contre (p. 8).

MAIN-D'OEUVRE

Nombre de personnes ayant travaillé le 15 mai 1985 dans l'exploitation (y compris les personnes qui en ont été empêchées par suite de circonstances imprévues, mais sauf celles occupées exclusivement aux soins du ménage) et à l'exclusion de la main-d'oeuvre utilisée pour des travaux à l'entreprise.

Personnes salariées et non salariées occupées aux cultures ou aux animaux de façon

	permanente (15)				non permanente			
	Code	H	Code	F	Code	H	Code	F
Chef d'exploitation (16)	400		401		416		417	
Conjoint(e) aidant le chef d'exploitation	402		403		418		419	
Autres membres de la famille du chef d'exploitation :								
salariés	404		405		420		421	
aidants (non salariés)	406		407		422		423	
Autres personnes :					424		425	
logées (17) et non nourries (18)	408		409					
nourries (18) et non logées (17)	410		411					
logées (17) et nourries (18)	412		413					
ni logées (17) ni nourries (18)	414		415					

ACCIDENTS DE TRAVAIL

Entre le 15 mai 1984 et le 15 mai 1985 y a-t-il eu un ACCIDENT DE TRAVAIL dans votre exploitation ?

	Code		
oui	286	1	☐
non		2	☐

Si oui, prière de remplir le questionnaire spécial prévu à cet effet, modèle IV.

SUCCESSION DU CHEF D'EXPLOITATION S'IL EST AGE DE 50 ANS ET PLUS

Y a-t-il, pour votre exploitation, un successeur présumé âgé de 14 ans et plus ?

	Code		
oui	287	1	☐
non		2	☐
vous ne savez pas		3	☐

Certifié exact et sincère, le.....1985

L'agent recenseur Le déclarant

Réservé à l'I.N.S. 999 .....

INSTRUCTIONS - QUESTIONNAIRE MOD. I.

(1) Triticale : Nouvelle variété céréalière basée sur un croisement de froment (triticum) avec du seigle (secale).

(2) Superficies déjà utilisées antérieurement comme superficies agricoles, mais qui ne sont plus utilisées à des fins agricoles pour des raisons économiques, sociales ou autres et qui n'entrent pas dans l'assolement. Ces superficies peuvent être utilisées de nouveau à l'aide de moyens normalement disponibles dans une exploitation (voir code 261, p.5)

(3) Sol des bâtiments, cours, chemins, étangs, jardins d'agrément, parcs, pelouses, terres incultes, etc... (voir code 263, p.5).

(4) Terres qui sont la propriété du responsable juridique et économique, y compris les terres mises en valeur à titre d'usufruitier, d'emphytéote ou d'autres titres équivalents. Les terres mises à la disposition d'un travailleur agricole comme avantage en nature comptent comme partie intégrante de l'exploitation qui met ces terres à disposition pour autant que le travailleur agricole n'utilise pas ses propres moyens de production.
Le lopin de terre d'un ascendant est considéré également comme partie intégrante de l'exploitation pour autant que les ascendants n'utilisent pas leurs propres moyens de production. (voir code 271, p.5).

(5) Terres louées par l'exploitation pour une redevance fixée d'avance (payée en espèces et/ou en nature) moyennant un contrat (écrit ou oral) de location.
Les terres louées par le fermier et mises à la disposition d'un travailleur agricole comme avantage en nature sont à comprendre dans l'exploitation considérée (voir code 272, p.5).

(6) Métayage : Terres exploitées en association par le propriétaire et le métayer sur la base d'un bail à métayage (écrit ou oral). Chacun des associés apporte à l'exploitation des facteurs de production et le produit est partagé entre eux selon une répartition convenue (voir code 273, p.5).

(7) RACES BOVINES RECONNUES

1. La RACE BLANC-BLEU est aussi connue sous le nom de race de Moyenne et Haute Belgique. Elle peut être mixte ou à viande mais doit dans les deux cas être renseignée sous le code 321, p.6. On la trouve surtout en Wallonie mais on peut la rencontrer dans les provinces flamandes.

2. La RACE PIE-ROUGE est blanche et rouge et surtout présente dans les provinces d'Anvers et du Limbourg, dans le Nord du Brabant et dans les cantons de l'Est. Certaines souches sont importées d'Allemagne. On trouve également dans le Luxembourg et dans d'autres provinces wallonnes (voir code 322, p.6).

3. La RACE PIE-NOIRE peut être belge, hollandaise, allemande ou américaine (Holstein, Friesian). C'est la race du pays de Herve et des Polders; mais on la trouve dans tout le pays (voir code 323, p.6). Dans le cas d'origine américaine, le bétail sera déclaré sous le code 328, p.6 (HOLSTEIN).

4. La RACE ROUGE se trouve en Flandre occidentale. On l'appelle aussi la race flamande. Elle est uniformément rouge, mais présente parfois des taches blanches (voir code 324, p.6).

5. La RACE BLANC-ROUGE se trouve en Flandre orientale. Elle est le plus généralement de corps blanc, de cou et tête rouges (voir code 325, p.6).

(8) DISTINCTION HYBRIDES - AUTRES RACES PORCINES.

HYBRIDES : Truies provenant d'une firme privée, spécialisée en élevage, reconnue ou non par l'état et, appliquant un programme raisonné de sélection et/ou métissage (voir code 393, p.6).

AUTRES RACES (voir code 394, p.6).
- Truies issues de tous autres croisements.
- Truies de races étrangères.

(9) Véhicules à moteur, à un essieu, utilisés en agriculture et horticulture, à l'exclusion des machines utilisées uniquement pour les parcs et les pelouses.

(10) A l'exclusion des tracteurs équipés d'outils tels que pelles, fourches, grues, etc. . . .

(11) Silos pourvus de parois latérales; ils peuvent être au niveau du sol ou partiellement enterrés.

(12) Silos dépourvus de parois latérales et constitués d'un tas de fourrage ensilé sur un sol durci.

(13) Machines (à couteau, à bêches, à chaîne ou à fraise) destinées à prélever le fourrage ensilé en silos horizontaux.

(14) Des explications détaillées et photos concernant les types de bâtiments d'élevage sont reprises dans la note spéciale aux agents recenseurs.

(15) Main-d'oeuvre permanente : toute personne qui travaille régulièrement à l'exploitation et dont l'activité totale atteint au moins la moitié du temps de travail que comporte l'année.

(16) Chef d'exploitation : n'indiquer qu'une seule personne par exploitation (une seule des rubriques 400 ou 401 ou 416 ou 417).

(17) Logées : dont le logement est assuré par le chef d'exploitation. Ce logement peut se trouver en dehors ou sur le lieu même de l'exploitation.

(18) Nourries : qui reçoivement au moins un repas principal par jour.

Informations

L'Institut national de statistique

Où trouver l'information de l'INS ?

Dans cinq grandes villes du pays, l'INS met à la disposition du public :

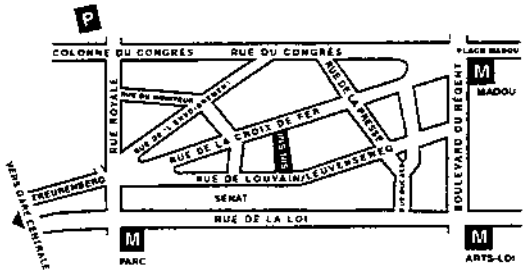
- ◊ Des annuaires et des publications spécialisées ainsi qu'une sélection de disquettes et de cédéroms.
- ◊ Une salle de lecture où il est possible de consulter les publications de l'INS, d'autres ministères ou d'institutions belges et internationales.

Toutes nos bibliothèques sont accessibles les jours ouvrables de 9h à 12h et de 13h à 16h.

Bruxelles

Rue de Louvain 44, 1000 Bruxelles
tél. 02/548.63.65 – 02/548.63.66 fax 02/548.63.67

Train (B) : Gare Centrale ou Congrès
Métro (M) : ligne 1A et 1B, station Parc ou Arts-Loi
ligne 2, station Arts-Loi ou Madou
Tram-Bus : 92, 93, 94 arrêt Parc
65, 66 arrêt Presse
29, 63 arrêt Congrès
Parking (P) : Cité administrative (payant)



Anvers

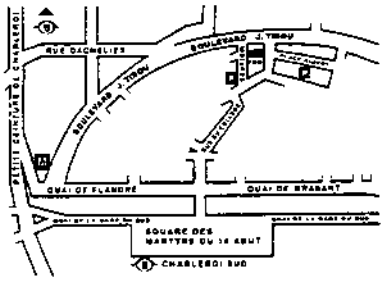
Italiëlei 124 - bus 85, 2000 Antwerpen
tél. 03/229.07.00 fax 03/233.28.30

Train (B) : Centraal Station
Métro (M) : arrêt Opera
Tram-Bus : accès facile (Fr. Rooseveltplaats)

Charleroi

Centre Albert, place Albert 1er, 8e étage, 6000 Charleroi
tél. 071/32.44.72 fax 071/32.44.71

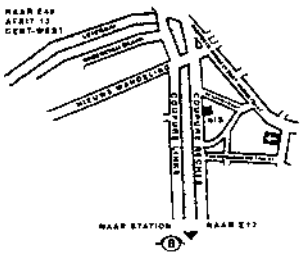
Train (B) : Charleroi Sud, 10 min depuis la gare (Place Buisset, Rue du Collège, Rue de Marchienne)
Bus : arrêt Tirou
Autoroute : petite ceinture de Charleroi - sortie Gare du Sud
Parking (P) : payant face à l'INS



Gand

Coupure rechts 620, 9000 Gent
tél. 09/225.77.16 fax 09/233.41.93

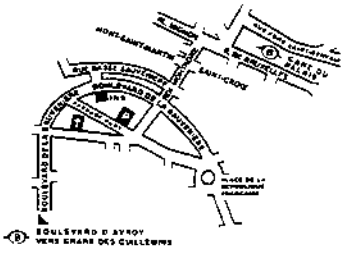
Train (B): Gent St. Pieters
Tram-Bus : 40, 43 arrêt Theresianenstraat
Autoroute: accès aisé par autoroute E40 (sortie N° 13 - Gent - West/Drongen)
Parking (P): au long de la "Coupure Rechts"



Liège

Bd de la Sauvenière 73-75, 4020 Liège
tél. 04/222.47.00 fax 04/222.49.94

Train (B): Gare des Guillemins ou Gare du Palais
Tram-Bus : (Guillemins) 1 et 4 arrêt Sauvenière
Parking (P): Neujean (à 20 m - même trottoir)
Mercure (en face)



L'INS diffuse de nombreux produits qui donnent une image chiffrée de la réalité socio-économique belge.

Ces produits, repris dans notre catalogue, sont disponibles auprès de nos centres régionaux ou auprès de notre service de Documentation - vente de Bruxelles. Notre catalogue vous sera envoyé sur simple demande. (voir adresses ci-contre).

Vous trouverez également un extrait de nos données, ainsi que la liste de nos publications sur notre site Internet : **http://statbel.fgov.be**

Publications analytiques

Études statistiques

La série «Études statistiques» présente des analyses couvrant les différents aspects de la situation de notre pays (territoire, démographie, société, économie, finances,...).
Nos statisticiens et d'autres auteurs vous y proposent des descriptions de méthodologie ou leurs réflexions basées sur nos statistiques.

Monographies du recensement

Des professeurs, des chercheurs et autres experts se sont penchés sur les résultats du Recensement de la Population et des Logements de 1991. Ces données exhaustives leur ont permis d'analyser en profondeur différents aspects de la situation belge face à l'évolution de la population et des familles, la fécondité, la mobilité, l'emploi, la scolarisation, le logement et l'urbanisation. Le fruit de leur collaboration est édité par l'INS dans une série de onze monographies disponibles en français et en néerlandais

Quelques autres publications

Publications générales

Annuaire statistique de la Belgique
Annuaire de statistiques régionales
Annuaire de poche
Bulletin de statistique - *Mensuelle*

Territoire et environnement

Statistique de l'occupation du sol (**disquette**)

Population

Statistiques démographiques - *Annuelle*

Société

Enquête sur les budgets des ménages - *Annuelle*

Économie et finances

Vente de biens immobiliers - *Annuelle*

Agriculture

Recensement agricole et horticole
au 15 mai - *Annuelle*

Industrie

Statistiques industrielles - *Mensuelle*

Commerce, services et transports

Statistique du commerce intérieur et des transports - *Mensuelle*

Études statistiques déjà publiées

Numéro	Titre
-	Tableau "Entrées-Sorties" de la Belgique pour 1959 (3 tomes), description générale de la méthode de calcul, demande finale au prix d'acquisition et investissements par branche d'activité, les coefficients techniques et la matrice inverse.
1	Analyse de la demande d'après les enquêtes sur les budgets des ménages effectués en Belgique en 1948-1949 et 1956-1957.
2	Croissance du revenu national de 1948 à 1959 et prévisions sur cette base pour les années à venir. Les dépenses des ménages en combustibles solides, électricité et gaz de ville de 1948 à 1959. Les élasticités de la demande des ménages en charbon, gaz et électricité aux prix et aux revenus d'après les séries chronologiques 1948-1959.-Prévisions relatives à la consommation des ménages en 1965.
3	Sur quelques aspects de la précision d'estimations basées sur les enquêtes de budgets ménagers. Répartition par province et par région linguistique du produit intérieur global et de la valeur ajoutée relative aux diverses branches d'activité.
4	Les comptes nationaux de la Belgique 1953-1962.
5	Enquête sur les budgets des ménages 1961 - Description de la méthode - Revenu, consommation et épargne de dix groupes sociaux.
6	La valeur ajoutée par branche d'activité et par travailleur dans les différentes provinces et régions linguistiques de 1955 à 1959. Évolution de la concentration industrielle, variation du rendement, des rémunérations, de la valeur ajoutée et des investissements avec la dimension des établissements industriels.
7	Enquête sur les budgets des ménages 1961 - Structure du budget selon les charges familiales et selon les régions linguistiques - Étude du caractère représentatif de l'enquête sur les budgets des ménages.
8	Les comptes nationaux de la Belgique 1953-1963 - Principales caractéristiques de l'évolution.
9	Enquête sur les budgets des ménages 1961 - Structure du budget selon la classe d'importance des communes et selon la branche d'activité ou le chef de ménage est occupé - Structure du budget selon l'épargne positive ou négative des ménages.
10	La révision 1964 de l'indice de la production industrielle. Indice de la production de biens intermédiaires, de biens de consommation et de biens d'investissement. Décomposition des séries chronologiques en leurs composantes suivant diverses méthodes - Application à quelques séries belges.
11	Les comptes nationaux de la Belgique 1953-1964 - Le développement économique et social.
12	Croissance économique des provinces et régions linguistiques 1955-1963.
13	Les comptes nationaux de la Belgique 1953-1965.
14	Situation actuelle de la statistique régionale. Orientation à l'exportation des différentes provinces et régions linguistiques. Répartition régionale du revenu national en 1961. Croissance économique des provinces et des régions linguistiques de 1962 à 1964.
15	Emploi et rémunération du travail par branche d'activité industrielle dans les provinces et régions linguistiques de 1955 à 1964.
16	Les comptes nationaux de la Belgique 1953-1966.
17	Typologie des communes belges d'après le degré d'urbanisation au 31 décembre 1961. Comparaison des enquêtes de 1961 et de 1963 sur les budgets des ménages d'ouvriers et d'employés.
18	Répartition de la valeur ajoutée des différentes branches d'activité et du produit intérieur global par province et par région linguistique - Années 1965 et 1966 Les indices régionaux de la production industrielle (base 1964=100); La réforme de l'indice des prix de détail.
19	Les comptes nationaux de la Belgique 1963-1967.
20	Les comptes nationaux de la Belgique 1965-1968.
21	Les comptes nationaux de la Belgique 1953-1969.
22	Tableau "Entrées-Sorties" de la Belgique pour 1965.
23	Croissance économique des provinces et régions linguistiques de 1965 à 1968. Orientation à l'exportation des différentes provinces et régions linguistiques - Années 1966 à 1968.
24	Vers un développement des comptes nationaux.
25	Les comptes nationaux de la Belgique 1966-1970.

Les publications de l'INS

Études statistiques déjà publiées

Numéro	Titre
26	Caractéristiques complémentaires de l'évolution économique selon les comptes nationaux 1963-1970. Les investissements des producteurs-distributeurs d'électricité: tests des hypothèses de l'accélération et de la capacité.
27	La division des communes belges en secteurs statistiques. Les investissements industriels des régions linguistiques de 1955 à 1969. Tableau "Entrées-Sorties" 1965.Données complémentaires sur l'emploi par branche d'activité.
28	Les comptes nationaux de la Belgique 1963-1971.
29	Les loyers des logements en 1970 et 1971.
30	Valeur ajoutée par travailleur dans l'industrie de 1953 à 1969. Les investissements industriels des provinces de 1955 à 1969.
31	Étude de quelques applications des équations de récurrence. Caractéristiques complémentaires de l'évolution économique selon les comptes nationaux 1963-1971.
32	Ajustements makehamiens, optimaux au sens des moindres carrés, d'une table de mortalité sur un intervalle d'âge déterminé. Croissance économique des provinces et régions linguistiques de 1966 à 1971.Valeur ajoutée et produit global par branche d'activité et région géographique.
33	Les comptes nationaux de la Belgique 1965-1972.
34	Les comptes nationaux de la Belgique. Estimations en prix de 1970 pour la période de 1953-1964. Caractéristiques complémentaires de l'évolution économique selon les comptes nationaux 1965-1972.
35	Tables de mortalité 1968-1972. A propos d'ajustements makehamiens d'une table de mortalité.
36	Les comptes nationaux de la Belgique 1966-1963.
37	L'industrie céramique de 1957 à 1972. Les carrières et les industries connexes de 1955 à 1972. L'industrie de la terre cuite de 1955 à 1972.
38	Enquête sur les budgets des ménages 1973-1974 (I).
39	Analyse des éléments actuariels résultant des tables de mortalité ajustées HS (1968-1972), HD (1968-1972) et HFR (1968-1972). Orientation à l'exportation des différentes provinces et régions linguistiques en 1971.Livraisons à l'étranger par branche d'activité et par région linguistique.
40	Les comptes nationaux de la Belgique 1966-1974.
41	Enquête sur les budgets des ménages 1973-1974 (II).
42	Tables de mortalité par régions linguistiques et pour la période 1968-1972. Quelques données de base sur l'évolution démographique de la Belgique et de ses régions. Le loyer des logements en 1973.
43	Caractéristiques complémentaires de l'évolution économique selon les comptes nationaux 1966-1974. Croissance économique des provinces et régions linguistiques de 1970 à 1974.Valeur ajoutée et produit global par branche d'activité et région géographique. Orientation à l'exportation des différentes provinces et régions linguistiques en 1972.Livraisons à l'étranger par branche d'activité industrielle et par région linguistique. Les investissements industriels des provinces et des régions linguistiques de 1970 à 1974.
44	Le chiffre d'affaires des entreprises et leurs achats de biens d'investissement depuis le 1er janvier 1971, sur base des déclarations à la taxe sur la valeur ajoutée.
45	Les comptes nationaux de la Belgique 1966-1975.
46	Enquête sur les budgets des ménages 1973-1974 (III).
47	Estimations de la population 1976-1985 pour le Royaume et ses régions. Tables de mortalité par état matrimonial pour la période 1968-1973.
48	Caractéristiques complémentaires de l'évolution économique selon les comptes nationaux 1966-1975. Comparaison de diverses mesures de la concentration dans les secteurs industriels en Belgique. Taux de variation moyen d'une série chronologique sur un intervalle de temps donné - Taux idéal de Pesek - Application à la croissance économique et aux investissements industriels des régions linguistiques et du Royaume de 1970 à 1974.
49	Tableau "Entrées-Sorties" de la Belgique pour 1970.

Études statistiques déjà publiées

Numéro	Titre
50	Enquête sur les budgets des ménages 1974-1974 (IV). Enquête sur les budgets des ménages d'indépendants 1973-1974.
51	Orientation à l'exportation des différentes provinces et régions linguistiques en 1973.Livraisons à l'étranger par branche d'activité industrielle et par région linguistique. Typologie des communes belges d'après le degré d'urbanisation au 31 décembre 1970. Réflexions sur l'emploi optimal des agents chargés des contrôles des recensements agricoles et horticoles en Belgique. La répartition des revenus personnels en Belgique: analyse statique. Les budgets communaux 1972-1975.
52	Les comptes nationaux de la Belgique 1967-1976. Caractéristiques complémentaires de l'évolution économique selon les comptes nationaux 1967-1976.
53	Orientation à l'exportation des différentes provinces et régions en 1974.Livraisons à l'étranger par branche d'activité industrielle et par région. Concentration industrielle en économie ouverte: le cas de la Belgique.
54	Les comptes nationaux de la Belgique 1968-1977.
55	Orientation à l'exportation des différentes provinces et régions en 1975.Livraisons à l'étranger par branche d'activité industrielle et par région. Croissance économique des provinces et régions de 1970 à 1976.Valeur ajoutée et produit global par branche d'activité et région géographique.
56	Les régions urbaines belges. Caractéristiques complémentaires de l'évolution économique selon les comptes nationaux 1968-1977. Les comptes sociaux européens.
57	Les comptes nationaux de la Belgique 1970-1978.
58	Statistiques au service des entreprises - Description et méthodologie.
59	Nouvelles perspectives de population (1976-2000) pour la Belgique, ses régions et ses arrondissements.
60	Orientation à l'exportation des différentes provinces et régions en 1976.Livraisons à l'étranger par branche d'activité industrielle et par région. Croissance économique des provinces et régions - Année 1977.Valeur ajoutée et produit global par branche d'activité et par région géographique.
61	Les comptes nationaux de la Belgique 1970-1979.
62	Orientation à l'exportation des différentes provinces et régions en 1977.Livraisons à l'étranger par branche d'activité industrielle et par région. Caractéristiques complémentaires de l'évolution économique selon les comptes nationaux.
63	Neuf cent mille étrangers en Belgique. Reflet de la récente évolution socio-économique et de la situation géographique du pays. Dispersion et relations de niveau élémentaire des noyaux d'habitat en Belgique. Situation en 1980 (avec carte hors texte).
64	Les comptes nationaux de la Belgique 1970-1980.
65	L'utilisation du sol en Belgique et son évolution depuis 1934 sur base des données cadastrales.
66	Caractéristiques complémentaires de l'évolution économique selon les comptes nationaux 1970-1980. Orientation à l'exportation des différentes provinces et régions en 1978.Livraisons à l'étranger par branche d'activité industrielle et par région.
67	Les comptes nationaux de la Belgique 1970-1981.
68	Localisation et structure de la population agricole belge.
69	Mortalité due au cancer en Belgique 1960-1979.Première analyse. Évolution de la concentration industrielle en Belgique. Croissance économique des provinces et régions de 1975 à 1979.Valeur ajoutée et produit global par branche d'activité et région géographique.
70	Tableau "Entrées-Sorties" de la Belgique pour 1975.
71	Les comptes nationaux de la Belgique 1971-1982.
72	Orientation à l'exportation des différentes provinces et régions en 1979.Livraisons à l'étranger par branche d'activité industrielle et par région. Subdivision des communes en quartiers en vue du recensement général de la population et des logements de 1981.
73	Les comptes nationaux de la Belgique 1970-1983.

Les publications de l'INS

Études statistiques déjà publiées

Numéro	Titre
74	Répartition régionale de l'activité de l'industrie des fabrications métalliques en 1982 et 1983 selon la statistique mensuelle. Quelques résultats du recensement général des logements du 1er mars 1981.
75	L'enquête de contrôle, un instrument permettant de déceler les erreurs dans les déclarations des effectifs et de déterminer la variance des réponses. Recensement agricole et horticole du 15 mai 1979 en Belgique.
76	Les comptes nationaux de la Belgique 1975-1984.
77	Quel avenir pour le recensement de la population? Le commerce extérieur de l'U.E.B.L en 1984.
78	Les comptes nationaux de la Belgique 1976-1985.
79	Croissance économique des provinces et régions de 1975 à 1984.
80	Quelques données sur l'évolution de la population active de 1970 à 1981.
81	Le commerce extérieur de l'U.E.B.L en 1986. Répartition régionale de l'activité des fabrications métalliques en 1984 et 1985 selon la statistique mensuelle.
82	Orientation à l'exportation des différentes provinces et régions, période 1980-1983.Livraisons à l'étranger par branche d'activité industrielle et par région.
83	Les comptes nationaux de la Belgique 1976-1986.
84	Tableau "Entrées-Sorties" de la Belgique pour 1980.
85	Les comptes nationaux de la Belgique 1976-1987.
86	Quelques données issues du recensement de la population au 1er mars 1981 sur la mobilité géographique de la main-d'oeuvre. Orientation à l'exportation des différentes provinces et régions. Période 1984-1985.Livraisons à l'étranger par branche d'activité industrielle et par région. Répartition régionale de l'activité de l'industrie des fabrications métalliques en 1986 et 1987 selon la statistique mensuelle.
87	Tableau "Entrées-Sorties" énergie de la Belgique pour 1980.
88	Les comptes nationaux de la Belgique 1980-1988.
89	Les régions urbaines belges en 1981. Orientation à l'exportation des différentes provinces et régions - Année 1986. Livraisons à l'étranger par branche d'activité industrielle et par région.
90	Les comptes nationaux de la Belgique 1980-1989.
91	Croissance économique des province et régions de 1980 à 1988. Orientation à l'exportation des différentes provinces et régions - Année 1987. Livraisons à l'étranger par branche d'activité industrielle et par région.
92	Les étrangers en Belgique d'après les recensements.
93	Les comptes nationaux de la Belgique 1980-1990.
94	La mortalité en Belgique à l'aube des années 90. Orientation à l'exportation des différentes provinces et régions - Année 1988. Livraisons à l'étranger par branche d'activité industrielle et par région.
95	Les comptes nationaux -SEC- 1980-1991. Agrégats - Comptes.
96	Les comptes nationaux -SEC- 1970-1991. Comptes et tableaux détaillés.
97	Les comptes nationaux -SEC- 1980-1992.
98	Les comptes nationaux -SEC- 1980-1992. Comptes et tableaux détaillés.
99	Orientation à l'exportation du Royaume, des régions, des provinces, des arrondissements et du Brabant flamand et wallon - Période 1980-1989. Livraisons industrielles à l'étranger par branche d'activité et par région.
100	Intrastat: une prévision des répercussions sur la qualité de la statistique belge du commerce intra-communautaire après 1992.
101	Les comptes nationaux -SEC- 1981-1993. Agrégats - Comptes.

Études statistiques déjà publiées

Numéro	Titre
102	Répartition régionale de l'emploi intérieur - Période 1980-1992.
103	Comptes nationaux -SEC- 1981-1993. Comptes et tableaux détaillés.
104	Les régions urbaines belges en 1991.
105	La fécondité en Belgique de 1991 à 1995.

Première édition

Achevé d'imprimer
par l'imprimerie de l'INS
B-1000 Bruxelles

ISSN 0069-8075, Novembre 1999